Winter 2010

FRONTIERS OF ENGINEERING

The BRIDGE

LINKING ENGINEERING AND SOCIETY

NATIONAL ACADEMY OF ENGINEERING
OF THE NATIONAL ACADEMIES

Promoting the technological welfare of the nation by marshalling the knowledge and insights of eminent members of the engineering profession.

The
BRIDGE

NATIONAL ACADEMY OF ENGINEERING

Irwin M. Jacobs, *Chair*
Charles M. Vest, *President*
Maxine L. Savitz, *Vice President*
Thomas F. Budinger, *Home Secretary*
George Bugliarello, *Foreign Secretary*
C.D. (Dan) Mote Jr., *Treasurer*

Editor in Chief (interim): George Bugliarello

Managing Editor: Carol R. Arenberg

Production Assistant: Penelope Gibbs

The Bridge (ISSN 0737-6278) is published quarterly by the National Academy of Engineering, 2101 Constitution Avenue, N.W., Washington, DC 20418. Periodicals postage paid at Washington, DC.

Vol. 40, No. 4, Winter 2010

Postmaster: Send address changes to *The Bridge*, 2101 Constitution Avenue, N.W., Washington, DC 20418.

Papers are presented in *The Bridge* on the basis of general interest and timeliness. They reflect the views of the authors and not necessarily the position of the National Academy of Engineering.

The Bridge is printed on recycled paper. ♻

© 2010 by the National Academy of Sciences. All rights reserved.

A complete copy of *The Bridge* is available in PDF format at *http://www.nae.edu/TheBridge*. Some of the articles in this issue are also available as HTML documents and may contain links to related sources of information, multimedia files, or other content.

The BRIDGE
Volume 40, Number 4 • Winter 2010
LINKING ENGINEERING AND SOCIETY

Editor's Note

Features

(continued on next page)

NAE News and Notes

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

Editor's Note

Andrew Weiner

Engineering on the Cutting Edge

Every year the NAE U.S. Frontiers of Engineering (FOE) Symposium brings together approximately 100 outstanding engineers, ages 30 to 45, to share ideas and learn about cutting-edge engineering research. A unique feature of FOE symposia is that participants are competitively selected from a wide range of engineering disciplines for the purpose of identifying individuals who are emerging as (or already may be) engineering leaders. These young, mid-career engineers are drawn from academic, industrial, and government institutions. FOE provides a unique opportunity for them to learn about frontiers in engineering areas other than their own and to network with promising young engineers in other fields.

The sixteenth U.S. FOE Symposium was held September 23–35, 2010, at the IBM Learning Center in Armonk, New York. The meeting was organized into four sessions with the following themes: cloud computing; engineering and music; autonomous aerospace systems; and engineering inspired by biology. Seven papers based on the symposium presentations are included in this issue of *The Bridge*.

Computer-based simulations and applications are now considered a "third-pillar" of scientific discovery complementing the traditional pillars of theory and experimentation. The first session, chaired by Dilma Da Silva of IBM Research and Ali Butt of Virginia Tech, focused on "Cloud Computing," a new paradigm in computing that lets users flexibly purchase resources from large-scale computing warehouses. Two papers on cloud computing are included in this issue of *The Bridge*.

The first paper, by Armando Fox, adjunct professor at UC-Berkeley, describes how cloud computing decouples lower level computer system details—including the need to purchase and administer one's own supercomputer or computer cluster—from application development, thus freeing users to focus on their technical/scientific missions. Dr. Fox also provides a vision of next-generation clouds and the main approaches being investigated by computer scientists to ensure that future clouds are amenable to wide-scale use and adaptation.

In the second paper, "Green Clouds: The Next Frontier," Parthasarathy Ranganathan, of Hewlett Packard Research Labs, discusses the implications for energy consumption and environmental consequences of using hundreds of thousands of computing nodes at a central location. Dr. Ranganathan provides examples of interdisciplinary research topics from building architecture to software design that have the potential to reduce the carbon footprint of the supporting cloud infrastructure.

Presentations in the second session on "Engineering and Music," chaired by Daniel Ellis of Columbia University and Youngmoo Kim of Drexel University, focused on how modern digital technology is revolutionizing the storage, distribution, retrieval, and creation of music—from an engineering perspective and an artisitic perspective. For some, this topic may go beyond preconceived ideas about the boundaries of engineering, but participants responded enthusiastically to the presentations in this session. Many commented that they found them enjoyable, stimulating, and thought provoking.

One paper from this session, authored by Elaine Chew of the University of Southern California, is published here. Her article, "Demystifying Music and Its Performance," describes how mathematics is being used to analyze and understand musical structure and how mathematical representations are being incorporated into visualizations for live performance. She provides several examples to give readers a sense of the richness and scope of research at the intersection of engineering and musical performance.

"Autonomous Aerospace Systems," the third session, was chaired by Michel Ingham of the Jet Propulsion Laboratory and Jack Langelann of Penn State University. The presentations addressed aspects of autonomy

that will change robotic systems from mechanisms that operate at the level of controlled systems that can function for a few minutes without human intervention to systems that can function autonomously for days or weeks in poorly characterized, or even unknown, environments. Examples were drawn from autonomous aeronautical and space autonomous systems.

Two papers from this session are included here. Mark Campbell, of Cornell University, describes research on probabilistic models of the environment and more efficient integration of human operators in the control/planning loop to enable deeper levels of intelligence in autonomous systems. In the second paper, Ella Atkins, University of Michigan, discusses the formidable challenges associated with the safe and efficient integration of unmanned aircraft systems into the National Airspace System and the role of automation and autonomy in the deployment of the next-generation air transportation system.

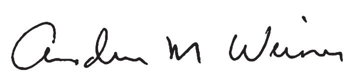
The final session, "Engineering Inspired by Biology," was chaired by Mark Byrne of Auburn University and Babak Parviz of the University of Washington. The presentations described the many ways biology is influencing interdisciplinary cutting-edge engineering research.

In the article by Jeffrey Fisher and Mostafa Ronaghi, of Illumina, a private company in San Diego, the authors discuss the status and future outlook for genomic, or DNA sequencing, technologies. The article focuses on prospects for increasing throughput and reducing cost, which will be crucial to opening new markets and applications, ranging from diagnosis to drug development.

In the final paper, Henry Hess, of Columbia University, reviews biologically inspired engineering on the molecular scale, including the design of active nanosystems that incorporate biomolecular motors and the study of self-assembly. He describes research that builds on nanotechnology and cell biology to replicate critical cellular functions to test our understanding of biological mechanisms and components.

In addition to the presentations, the symposium included extended, lively Q&A sessions, panel discussions, and poster sessions and other activities that encourage personal discussions and networking. The dinner speaker this year, a traditional highlight of FOE programs, was Dr. Bernard Meyerson, IBM Fellow, IBM Vice President for Innovation, and a pioneer and leader in high-performance semiconductor technologies. In his inspiring talk, "Radical Innovation to Create a Smarter Planet," Dr. Meyerson described innovation in large organizations. He provided examples of how radical innovation, integrated with radical collaboration, was enabling IBM to keep ahead while creating new opportunities for contributions to sustainability and improving people's well-being. Dr. Meyerson was an enthusiastic presence throughout the three-day conference, and he and his IBM colleagues were generous and attentive hosts.

It has been my great privilege to serve again as chair of the Organizing Committee for the 2010 U.S. FOE Symposium. I would like to express my gratitude to Janet Hunziker, NAE Senior Program Officer, and Lance Davis, NAE Executive Officer, for their contributions to the planning and implementation of this unique meeting and to thank the sponsors of the symposium—IBM, The Grainger Foundation, the Air Force Office of Scientific Research, the Defense Advanced Research Projects Agency, the U.S. Department of Defense (Office of the Director, Defense Research and Engineering), the National Science Foundation, Microsoft Research, and Cummins Inc.



The essence of cloud computing is to make datacenter hardware and software available to the general public on a pay-as-you-go basis.

Opportunities and Challenges of Cloud Computing

Armando Fox

Armando Fox is adjunct professor, Reliable Adaptive Distributed Systems Laboratory (RAD Lab), University of California, Berkeley.

Computer science is moving forward so quickly and is so focused on its recent history that we are often surprised to learn that visionary ideas were articulated long before the technology for their practical implementation was developed. The following vision of "utility computing" is excerpted from an overview of the pioneering and highly influential MULTICS computing system (Corbató and Vyssotsky, 1965):

> One of the overall design goals is to create a computing system which is capable of meeting almost all of the present and near-future requirements of a large computer utility. Such systems must run continuously and reliably 7 days a week, 24 hours a day in a way similar to telephone or power systems, and must be capable of meeting wide service demands . . . [T]he importance of a multiple access system operated as a computer utility is that it allows a vast enlargement of the scope of computer-based activities, which should in turn stimulate a corresponding enrichment of many areas of our society.

Today, 45 years later, that vision appears close to becoming reality. In 2008, Amazon announced the availability of its Elastic Compute Cloud (EC2), making it possible for anyone with a credit card to use the servers in Amazon's datacenters for 10 cents per server hour with no minimum or maximum purchase and no contract (Amazon AWS, 2008b). Amazon has since added options and services and reduced the base price to 8.5 cents per server hour.) The user is charged for only as long as he/she uses the computer rounded up to the next hour.

The essence of cloud computing is making datacenter hardware and software available to the general public on a pay-as-you-go basis. Every user enjoys the illusion of having virtually infinite capacity available instantaneously on demand. Hence the term *utility computing* is used to describe the "product" sold by a cloud-computing provider.

TABLE 1 Comparative Economies of Scale in 2006 for a Medium-Sized Datacenter (≈1,000 servers) and a Warehouse-Scale Datacenter (≈50,000 servers).

Technology	Medium-Sized Data Center	Warehouse-Scale Data Center	Ratio
Network	$95 per Mbit/sec/month[a]	$13 per Mbit/sec/month	7.1
Storage	2.20 per GByte/month[b]	$0.40 per GByte/month	5.7
Administration	1 administrator per ≈140 servers	1 administrator for > 1,000 servers	7.1

[a]Mbit/sec/month = megabit per second per month. [b]GByte/month = gigabyte per month.
Source: Hamilton, 2008.

Of course, by 2008, many companies, such as Google Search and Microsoft Hotmail, were already operating extensive "private clouds" that delivered proprietary SaaS (software as a service). These companies had found it necessary to develop the programming and operational expertise to run such installations.

In contrast, EC2 was the first truly low-cost utility computing that was not bundled with a particular SaaS application. Users of EC2 were allowed to deploy applications of their choice, which greatly increased the popularity of the system. Private-cloud operators Google and Microsoft soon followed suit and now provide public-cloud services in addition to their proprietary services.

At first, skeptics were hard pressed to believe that Amazon could operate such a service at a profit. But, as leading software architect James Hamilton observed (2008), because of economies of scale, the costs of bandwidth, storage, and power for warehouse-scale datacenters are five to seven times cheaper than for medium-sized datacenters (see Table 1). With Amazon's retail-to-consumer operational expertise, the company found a profitable way to pass these savings along to individual users.

Cost Associativity and Elasticity

The cloud-computing service model, which represents a radical departure from conventional information technology (IT), enables *fundamentally new* kinds of computation that were previously infeasible. For example, in 2008, the National Archives released 17,481 pages of documents, including First Lady Hillary Clinton's daily schedule of activities. Peter Harkins, a senior engineer at *The Washington Post*, using 200 computers in EC2 for less than nine hours, produced a searchable corpus of the documents and made it publicly available on the World Wide Web less than a day later (Amazon AWS, 2008b). The server time cost Harkins less than $150—the same cost as using a single server for 1,800 hours, and far less than the cost of purchasing a single server outright. Being able to use 200 servers for nine hours for the same price as using one server for 1,800 hours is an unprecedented new capability in IT that can be called *cost associativity*.

That same year, 2008, programmers at the Web start-up company Animoto developed an application to create music videos from a user's photo collection. When that application was made available to the more than 200 million users of Facebook, it became so popular so quickly that the number of users doubled every 12 hours for the next three days, causing the number of servers to increase from 50 to 3,500. After the peak subsided, demand fell to a much lower level, and the unnecessary servers were released.

Elasticity, the ability to add and remove servers in minutes, rather than days or weeks, is also unprecedented in IT. Elasticity is financially appealing because it allows actual usage to closely track demand on an hour-by-hour basis, thereby transferring the *risk* of making a poor provisioning decision from the service operator to the cloud-computing provider.

But elasticity is even more important for handling *spikes* and *data hot spots* resulting from unexpected events. During the terrorist attacks of September 11, 2001, for example, viewer traffic on the CNN website increased by an order of magnitude in just 15 minutes (LeFebvre, 2001). In another case, when entertainer Michael Jackson died unexpectedly in 2009, the number of Web searches about Jackson spiked to nearly 10 times the average so suddenly that Google initially mistook

the event for a malicious attack on its search service.

According to Tim O'Reilly, founding editor of O'Reilly Media, a leading technical publisher, the ability to deal with sudden surges is particularly important for mobile applications that "respond in real time to information provided either by their users or by nonhuman sensors" (quoted in Siegele, 2008). In other words, these services are accessible to the more than 50 percent of the world population equipped with cell phones, the most ubiquitous Internet access devices.

Opportunities and Challenges

Scaling Down

Before the advent of cloud computing, scaling up was considered a permanent change, because it usually meant buying and installing new hardware. Consequently, extensive research was conducted on scaling up systems without taking them offline. The idea of subsequently scaling them down—and then possibly back up again—was not even considered.

Since cloud computing involves borrowing machines from a shared pool that is constantly upgraded, scale-up and scale-down are likely to mean that hardware will be more heterogeneous than in a conventional datacenter. Research is just beginning on software, such as scalable consistency-adjustable data storage (SCADS), which can gracefully scale down as well as up in a short time (Armbrust et al., 2009).

At the other extreme, fine-grained pricing may enable even cheaper utility computing during demand troughs. California power companies have already introduced demand-based pricing models in which power is discounted during off-peak times. By analogy, Amazon EC2 has introduced a new mechanism whereby otherwise unused machines are made available at a discounted rate on a "best-effort" basis. However, the user might be forced to give up the machine on short notice if demand increases and a priority customer is willing to pay a premium for it.

This leads to a relatively new situation of clusters whose topologies and sizes can change at any time and whose cycles may be "reclaimed" on short notice for higher priority applications. Research on scheduling frameworks, such as Mesos, is addressing how applications on cloud computing can deal gracefully with such fluctuations (Hindman et al., 2010).

The ability to scale down also introduces new motivations for improving the energy efficiency of IT. In traditional research proposals, energy costs are usually absorbed into general institutional overhead. With cloud computing, a customer who uses fewer machines consumes less energy and, therefore, pays less. Although warehouse-scale datacenters are now being built in locations where cheaper power (e.g., hydroelectric power) is available (Table 2), the pay-as-you-go model of cloud computing introduces a direct financial incentive for cloud users to reduce their energy usage.

Several challenges, however, may interfere with this opportunity for "greener" IT. Unfortunately, today's servers consume nearly half as much energy when they are idle as when they are used. Barroso and Hölzle (2007) have argued that we will need design improvements at all levels, from the power supply to energy-aware software, to achieve "energy proportional" computing in which the amount of energy consumed by a server is proportional to how much work it does.

Better and Faster Research

Cost associativity means that "embarrassingly parallel" experiments—experiments that require many trials or tasks that can be pursued independently—can be accelerated to the extent that available cloud resources allow. For example, an experiment that requires 100,000 trials of one minute each would take more than two months to complete on a single server. Cost associativity makes it possible to harness 1,000 cloud servers for two hours for the same cost. Researchers in the RAD Lab working on datacenter scale computing now routinely run experiments involving hundreds of servers to test out their ideas at realistic scale. Before cloud computing, this was impossible for any university laboratory.

TABLE 2 Price of Kilowatt Hours (kWh) of Electricity

Cents per kWh	Region	Factors
3.6	Idaho	Hydroelectric power; no long-distance transmission
10.0	California	Long-distance transmission; limited transmission lines in Bay Area; no coal-fired electricity allowed in the state
18.0	Hawaii	Fuel must be shipped to generate electricity

Source: EIA, 2010.

Tools like Google's MapReduce (Dean and Ghemawat, 2004) and the open-source equivalent, Hadoop, give programmers a familiar data-parallel "building block" and encapsulate the complex software engineering necessary for handling the challenges of resource scheduling and responding to machine failures in the cloud environment. However, because many problems cannot be easily expressed as MapReduce tasks, other frameworks, such as Pig, Hive, and Cascading, have emerged that provide higher level languages and abstractions for cloud programming.

Indeed, Amazon's recently-introduced "Elastic MapReduce" service, which provides a "turnkey" version of the MapReduce framework, allows jobs to be written using not only those frameworks, but also statistical modeling packages, such as R. On the level of cloud infrastructure itself, the goal of the Berkeley BOOM project (*boom.cs.berkeley.edu*) is to simplify the creation of new cloud programming frameworks by applying principles from declarative networking.

Progress is being made on all of these fronts, and some new systems are in regular use in production environments. However, the artifacts and ecosystem comprising them are still a long way from "turnkey" systems that will allow domain-expert programmers to seamlessly combine the abstractions in their applications.

University laboratories can now run experiments involving hundreds of servers.

High-Performance Computing

The scientific and high-performance computing (HPC) community has recently become more interested in cloud computing. Compared to SaaS workloads, which rely on request-level parallelism, HPC workloads typically rely on thread- or task-level parallelism, making them more communication-intensive and more sensitive to communication latency. These properties make HPC workloads particularly vulnerable to "performance noise" artifacts introduced by the pervasive use of virtualization in cloud environments (Armbrust et al., 2010b).

Legacy scientific codes often rely on resource-scheduling approaches, such as gang scheduling and make assumptions about the network topology that connects the servers. Such design decisions make sense in a statically provisioned environment but not for cloud computing. Thus, not surprisingly, early benchmarks of existing HPC applications on public clouds were not encouraging (Evangelinos and Hill, 2008; Walker, 2008).

However, cloud providers have been quick to respond to the potential HPC market, as illustrated by Amazon's introduction in July 2010 of "Cluster Compute Instances" tuned specifically for HPC workloads. Experiments at the National Energy Research Scientific Computing (NERSC) Laboratory at Lawrence Berkeley Laboratory measured an 8.5X performance improvement on several HPC benchmarks when using this new type of instance compared to conventional EC2 instances. Amazon's own measurements show that a "virtual cluster" of 880 HPC instances can run the LINPACK linear algebra benchmark faster than the 145th-fastest supercomputer in the world, as measured by Top500.com. These results have encouraged more scientists and engineers to try cloud computing for their experiments. Installations operated by academic/industrial consortia, such as the Google/IBM/NSF CluE cluster that runs Hadoop (NSF, 2009), Yahoo's M45 cluster (*http://labs.yahoo.com/Cloud_Computing*), and OpenCirrus (*opencirrus.org*), are other examples of cloud computing for scientific research.

Even if the running time of a problem is slower on cloud computing than on a dedicated supercomputer, the total time-to-answer might still be shorter with cloud computing, because unlike traditional HPC facilities, the user can provision a "virtual supercomputer" in the cloud instantly rather than waiting in line behind other users (Foster, 2009).

Longtime HPC veteran Dan Reed, now head of the eXtreme Computing Group (XCG) at Microsoft Research, also believes cloud computing is a "game changer" for HPC (West, 2009). He points out that while cloud infrastructure design shares many of the challenges of HPC supercomputer design, the much larger volume of the cloud infrastructure market will influence hardware design in a way that traditional HPC has been unable to do.

Transfers of Big Data

According to Wikipedia, the Large Hadron Collider could generate up to 15 petabytes (15×10^{15} bytes) of data per year, and researchers in astronomy, biology, and

many other fields routinely deal with multi-terabyte (TB) datasets. A boon of cloud computing is its ability to make available tremendous amounts of computation on-demand with large datasets. Indeed, Amazon is hosting large public datasets for free, perhaps hoping to attract users to purchase nearby cloud computing cycles (Amazon AWS, 2008a).

The key word here is *nearby*. Transferring 10 TB over a network connection at 20 megabits per second—a typical speed observed in measurements of long-haul bandwidth in and out of Amazon's S3 cloud storage service (Garfinkel, 2007)—would take more than 45 days and incur transfer charges of $100 to $150 per TB.

In the overview of cloud computing by Armbrust et al. (2010b), we therefore proposed a service that would enable users to instead ship crates of hard drives containing large datasets overnight to a cloud provider, who would physically incorporate them directly into the cloud infrastructure. This idea was based on experience with this method by the late Jim Gray, the Turing Award-winning computer scientist who was recently instrumental in promoting the use of large-scale computation in science and engineering. Gray reported using this technique reliably; even if disks are damaged in transit, well-known RAID-like techniques could be used to mitigate the effects of such failures (Patterson, 2003).

Shortly after the overview was published, Amazon began offering such a service and continues to do so. Because network cost/performance is improving more slowly than any other cloud computing technology (see Table 3), the "FedEx a disk" option for large data transfers is likely to become increasingly attractive.

TABLE 3 Update of Gray's Costs of Computing Resources from 2003 to 2008

	Wide-area (long-haul) Network Bandwidth/ Month	CPU Hours (all cores)	Disk Storage
Item in 2003	1 Mbps WAN[a] link	2 GHz CPU, 2 GB DRAM	200 GB disk, 50 Mb/s transfer rate
Cost in 2003	$100/month	$2,000	$200
What $1 buys in 2003	1 GB	8 CPU hours	1 GB
Item in 2008	100 Mbps WAN link	2 GHz, 2 sockets, 4 cores/socket, 4 GB DRAM	1 TB disk, 115 MB/s sustained transfer
Cost in 2008	$3,600/month	$1,000	$100
What $1 buys in 2008	2.7 GB	128 CPU hours	10 GB
Cost/ performance improvement	2.7x	16x	10x
Cost to rent	$0.27–$0.40	$2.56	$1.20–$1.50
What $1 buys on AWS[b] in 2008	$0.10–$0.15/ GB × 3 GB	128× 2 VMs@ $0.10 each	$0.12–$0.15/ GB-month × 10 GB

[a]WAN = wide-area (long-haul) network. [b]AWS = Amazon Web Services
Source: Armbrust et al., 2010a.

Licensing and Cloud Provider Lock-In

Amazon's EC2 represents one end of a spectrum in that its utility computing service consists of a bare-bones server built around the Intel x86 processor architecture. Cloud users must provide all of the software themselves, and open-source building blocks, such as the Linux operating system, are popular starting points. However, scientific and engineering research also frequently requires the use of proprietary software packages, such as Matlab.

Although some publishers of proprietary software (including Matlab) now offer a pay-as-you-go licensing model like the model used for the public cloud, most software is still licensed in a "cloud-unfriendly" manner (e.g., per seat or per computer). Changing the structure of software licenses to approximate the public cloud pricing model is a nontechnical but real obstacle to the increased use of the cloud in scientific computing.

In addition, if other providers, such as Google AppEngine or Microsoft Azure, provide value-added software functionality in their clouds, users might become dependent on such software to the point that their computing jobs come to require it. An example is Google AppEngine's automatic scale-up and scale-down functionality, which is available for certain kinds of user-deployed applications. If such applications were migrated to a non-Google platform, the application authors might have to create this functionality themselves.

The potential risk of "lock-in" to a single provider could be partially mitigated by standardizing the application programming interfaces and data formats used by different cloud services. Providers could then differentiate their offerings by the quality of their implementations, and migration from one provider to another would result in a possible loss of performance, rather than a loss of functionality. The Data Liberation Front, a project started by a group of Google engineers, is one group that is actively pursuing data standardization.

Conclusion

In 1995, researchers at Berkeley and elsewhere had argued that networks of commodity workstations (NOWs) offered potential advantages over high-performance symmetrical multiprocessors (Anderson et al., 1995). The advantages would include better scalability, cost-effectiveness, and potential high availability through inexpensive redundancy.

At that time software could not deal with important aspects of NOW architecture, such as the possibility of partial failure. Nevertheless, the economic and technical arguments for NOW seemed so compelling that, over the course of several years, academic researchers and commercial and open-source software authors developed tools and infrastructure for programming this idiosyncratic architecture at a much higher level of abstraction. As a result, applications that once took years for engineers to develop and deploy on a NOW can be prototyped today by Berkeley undergraduates as an eight-week course project.

Given this rapid evolution, there is good reason to be optimistic that in the near future computer-based scientific and engineering experiments that take weeks today will yield results in a matter of hours. When that time arrives, the necessity of purchasing and administering one's own supercomputer or computer cluster (and then waiting in line to use it) will seem as archaic as text-only interfaces do today.

Acknowledgments

This work was done at the University of California Berkeley Reliable Adaptive Distributed Systems Laboratory (RAD Lab), and much of what is reported here builds on a survey article written by Michael Armbrust, the present author, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. Support was provided by Sun Microsystems, Google, Microsoft, Amazon Web Services, Cisco Systems, Cloudera, eBay, Facebook, Fujitsu, Hewlett-Packard, Intel, Network Appliance, SAP, VMWare, and Yahoo! Inc., as well as matching funds from the State of California MICRO Program (Grants 06-152, 07-010, 06-148, 07-012, 06-146, 07-009, 06-147, 07-013, 06-149, 06-150, and 07-008), the National Science Foundation (Grant CNS-0509559), and the University of California Industry/University Cooperative Research Program (UC Discovery) Grant COM07-10240.

References

Amazon AWS. 2008a. Public Data Sets on AWS. Available online at *http://aws.amazon.com/publicdatasets/*.

Amazon AWS. 2008b. AWS Case Study: Washington Post. Available online at *http://aws.amazon.com/solutions/case-studies/washington-post*.

Anderson, T.E., D.E. Culler, and D. Patterson. 1995. A case for NOW (networks of workstations). IEEE Micro 15(1): 54–64.

Armbrust, M., A. Fox, D.A. Patterson, N. Lanham, B. Trushkowsky, J. Trutna, and H. Oh. 2009. SCADS: scale-independent storage for social computing applications. In CIDR Perspectives 2009. Available online at *http://www.db.cs.wisc.edu/cidr/cidr2009/Paper_86.pdf*.

Armbrust, M., A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. 2010a. Above the Clouds: A Berkeley View of Cloud Computing. Technical Report EECS-2009-28, EECS Department, University of California, Berkeley. Available online at *www.abovetheclouds.cs.berkeley.edu*.

Armbrust, M., A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. 2010b. A view of cloud computing. Communications of the ACM 53(4): 50–58.

Barroso, L.A., and U. Hölzle. 2007. The case for energy-proportional computing. IEEE Computer 40(12): 33–37.

Corbató, F.J., and V.A. Vyssotsky. 1965. Introduction and overview of the multics system. P. 185 in Proceedings of the Fall Joint Computer Conference, 1965. New York: IEEE.

Dean, J., and S. Ghemawat. 2004. MapReduce: Simplified data processing on large clusters. Pp. 137–150 in Proceedings of the 6th USENIX Symposium on Operating Systems Design and Implementation (OSDI '04), December 5–8, 2004, San Diego, Calif. Berkeley, Calif.: USENIX.

EIA (Energy Information Administration). 2010. State Electricity Prices, 2006. Available online at *http://www.eia.doe.gov/neic/rankings/stateelectricityprice.htm*.

Evangelinos, C., and C.N. Hill. 2008. Cloud computing for parallel scientific HPC applications: feasibility of running coupled atmosphere-ocean climate models on Amazon's EC2. In First ACM Workshop on Cloud Computing and its Applications (CCA'08), October 22–23, 2008, Chicago, Ill. New York: ACM.

Foster, I. 2009. What's faster—a supercomputer or EC2? Available online at *http://ianfoster.typepad.com/blog/2009/08/whats-fastera-supercomputer-or-ec2.html*.

Garfinkel, S. 2007. An Evaluation of Amazon's Grid Computing Services: EC2, S3 and SQS. Technical Report TR-08-07. Harvard University. Available online at *http://simson.net/clips/academic/2007.Harvard.S3.pdf*.

Hamilton, J. 2008. Internet-Scale Service Efficiency. Presentation at 2nd Large-Scale Distributed Systems and Middleware (LADIS) Workshop, September 15–17, 2008, White Plains, NY. Available online at *http://www.cs.cornell.edu/courses/cs5410/2008fa/Slides/MSCloud.pdf*.

Hindman, B., A. Konwinski, M. Zaharia, A. Ghodsi, A.D. Joseph, R.H. Katz, S. Shenker, and I. Stoica. 2010. Mesos: A Platform for Fine-Grained Resource Sharing in the Data Center. Technical Report UCB-EECS-2010-87. University of California, Berkeley. Available online at *http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-87.html*.

LeFebvre, W. 2001. CNN.com: Facing a World Crisis. In Proceedings of the 15th Conference on Systems Administration (LISA 2001), San Diego, Calif., December 2–7, 2001. Berkeley, Calif.: USENIX.

NSF (National Science Foundation). 2009. National Science Foundation Awards Millions to Fourteen Universities for Cloud Computing Research. Available online at *http://www.nsf.gov/news/news_images.jsp?cntn_id=114686&org=NSF*.

Patterson, D. 2003. A Conversation with Jim Gray. ACM Queue 4(1). Available online at *http://queue.acm.org/detail.cfm?id=864078*.

Siegele, L. 2008. A Survey of Corporate IT: Let It Rise. The Economist (October 23). Web edition only.

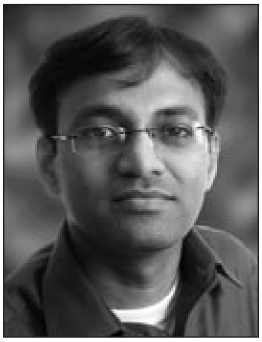
Walker, E. 2008. Benchmarking Amazon EC2 for high performance scientific computing. ;login 33(5): 18–23. Available online at *http://www.usenix.org/publications/login/2008-10/openpdfs/walker.pdf*.

West, J. 2009. Twins Separated at Birth: Cloud Computing, HPC and How Microsoft is Trying to Change How We Think About Scale. Available online at *http://www.hpcwire.com/features/Twins-Separated-at-Birth-41173917.html*.

New thinking about computer systems is focused on holistic designs that cross traditional boundaries.

Green Clouds

The Next Frontier



Parthasarathy Ranganathan is a Distinguished Technologist, Hewlett Packard Research Labs.

Parthasarathy Ranganathan

We are entering an exciting era for computer-systems design. In addition to continued advances in performance, next-generation designs are also addressing important challenges related to power, sustainability, manageability, reliability, and scalability. At the same time, new combinations of emerging technologies (e.g., photonics, non-volatile storage, and 3D stacking), and new workloads (related to cloud computing, unstructured data, and virtualization) are presenting us with new opportunities and challenges. The confluence of these trends has led us to rethink the way we design systems—motivating holistic designs that cross traditional design boundaries.

In this article, we examine what this new approach means for the basic building blocks of future systems and how to manage them. Focusing on representative examples from recent research, we discuss the potential for dramatic (10 to 100X) improvements in efficiency in future designs and the challenges and opportunities they pose for future research.

Predicting the Future of Computing Systems

What can we predict for computing systems 10 years from now? Historically, the first computer to achieve terascale computing (10^{12}, or one trillion computing operations per second) was demonstrated in the late 1990s. About 10 years later, in mid-2008, the first petascale computer was

demonstrated at 1,000 times more performance capability. Extrapolating these trends, one can expect an exascale computer by approximately 2018. That is a staggering *one million trillion* computing operations per second and a thousand-fold improvement in performance over any current computer.

Moore's law (often described as the trend that computing performance doubles every 18 to 24 months) has traditionally helped predict performance challenges, for terascale and more recently petascale computing, but the transition from petascale to exascale computing is likely to pose some new challenges we need to address going forward.

Challenges

The Power Wall

The first challenge is related to what is commonly referred to as the *power wall*. Power consumption is becoming a key constraint in the design of future systems. This problem is manifested in several ways: in the amount of electricity consumed by systems; in the ability to cool systems cost effectively; in reliability; and so on.

For example, recent reports indicate that the electricity costs for powering and cooling cloud datacenters can be millions of dollars per year, often more than was spent on buying the hardware (e.g., Barroso and Hözle, 2007)! IDC, an industry analyst firm, has estimated that worldwide investment in power and cooling was close to $40 billion last year (Patel, 2008).

This emphasis on power has begun to have a visible impact on the design of computing systems, as system design constraints are shifting from optimizing performance to optimizing energy efficiency or performance achieved per watt of power consumed in the system. This shift has been partly responsible for the emergence of multi-core computing as the dominant way to design microprocessors.

In addition, recognition has been growing that designers of energy-efficiency optimized systems must take into consideration not only power consumed by the computing system, but also power consumed by the supporting equipment. For example, for every watt of power consumed in the server of a datacenter, an additional half to one watt of power is consumed in the equipment responsible for power delivery and cooling (often referred to as the burdened costs of power and cooling, or power usage effectiveness [PUE] [Belady et al., 2008]).

Sustainability

Sustainability is also emerging as an important issue. The electricity consumption associated with information technology (IT) equipment is responsible for 2 percent of the total carbon emissions in the world, more than the emissions of the entire aviation industry. More important, IT is increasingly being used as the tool of choice to address the remaining 98 percent of carbon emissions from non-IT industries (e.g., the use of video conferencing to reduce the need for travel or the use of cloud services to avoid transportation or excess manufacturing costs) (Banerjee, 2009).

One way to improve sustainability is to consider the total life cycle of a system—including both the supply and demand side. In other words, in addition to the amount of energy used in *operating* a system, it is important to consider the amount of energy used in *making* the system.

The transition from petascale to exascale computing poses significant new challenges.

Manageability

Sustainability is just one of the new "*ilities*" that pose challenges for the future. Another key challenge pertains to *manageability*, which can be defined as the collective processes of deployment, configuration, optimization, and administration during the life cycle of an IT system.

To illustrate this challenge, consider, as an example, the potential infrastructure in a future cloud datacenter. On the basis of recent trends, one can assume that there will be five global datacenters with 40 modular containers each, 10 racks per container, 4 enclosures per rack, and 16 blade servers per enclosure. If each blade server has two sockets with 32 cores each and 10 virtual machines per core, this cloud vendor will have a total of 81,920,000 virtual servers to operate its services. Each of the more than 80 million servers, in turn, will require several classes of operations—for bring-up, day-to-day operations, diagnostics, tuning, and other processes, ultimately including retirement or redeployment of the system. Although a lot of work has been done on managing computer systems, manageability on such a large scale poses new challenges.

Reliability

Trends in technology scaling circuit level and increased on-chip integration at the micro-architectural level lead to a higher incidence of both transient and permanent errors. Consequently, new systems must be designed to operate reliably and provide continued up-time, even when they are built of unreliable components.

Business Trends

Finally, these challenges must be met within the constraints of recent business trends. One important trend is the emphasis on reducing total costs of ownership for computing solutions. This often translates to a design constraint requiring the use of high-volume commodity components and avoiding specialization limited to niche markets.

Opportunities

We believe that the combination of challenges—low power, sustainability, manageability, reliability, and costs—is likely to influence how we think about system design to achieve the next 1,000-fold increase in performance for the next decade. At the same time, we recognize that interesting opportunities are opening up as well.

The amount of online data is estimated to have increased 60-fold in the last seven years.

Data-Centric Workloads

A fundamental shift has taken place in terms of data-centric workloads. The amount of data being created is increasing exponentially, much faster than Moore's law predicted. For example, the size of the largest data warehouse in the Winter Top Ten Survey has been growing at a cumulative annual growth rate of 173 percent (Winter, 2008). The amount of online data is estimated to have increased nearly 60-fold in the last seven years, and data from richer sensors, digitization of offline content, and new applications like Twitter, Search, and others will surely increase data growth rates. Indeed, it is estimated that only 5 percent of the world's off-line data has been digitized or made available through online repositories so far (Mayer, 2009).

The emergence and rapid increase of data as a driving force in computing has led to a corresponding increase in data-centric workloads. These workloads focus on different aspects of the data life cycle (capture, classify, analyze, maintain, archive, and so on) and pose significant challenges for the computing, storage, and networking elements of future systems.

Among these, an important recent trend (closely coupled with the growth of large-scale Internet web services) has been the emergence of complex analysis on an immense scale. Traditional data-centric workloads like web serving and online transaction processing (e-commerce) are being superseded by workloads like real-time multimedia streaming and conversion; history-based recommendation systems; searches of texts, images, and even videos; and deep analysis of unstructured data (e.g., Google Squared).

Emerging data-centric workloads have changed our assumptions about system design. These workloads typically operate at larger scale (hundreds of thousands of servers) and on more diverse data (e.g., structured, unstructured, rich media) with input/output (I/O) intensive, often random data-access patterns and limited locality. Another characteristic of data-centric workloads is a great deal of innovation in the software stack to increase scalability and commodity hardware (e.g., Google MapReduce/BigTable).

Improvements in Throughput, Energy Efficiency, Bandwidth, and Memory Storage

Recent trends suggest several potential technology disruptions on the horizon (Jouppi and Xie, 2009). On the computing side, recent microprocessors have favored multi-core designs that emphasize multiple simpler cores for greater throughput. This approach is well matched with the large-scale distributed parallelism in data-centric workloads. Operating cores at near-threshold voltage has been shown to significantly improve energy efficiency. Similarly, recent advances in networking, particularly related to optics, show a strong growth in bandwidth for communication among computing elements at various levels of the system.

Significant changes are also expected in the memory/storage industry. Recently, new non-volatile RAM (NVRAM) memory technologies have been demonstrated that significantly reduce latency and improve energy efficiency compared to Flash and Hard Disk. Some of these NV memories, such as phase-change RAM (PCRAM) and Memristors, have shown the

potential to replace DRAM with competitive performance and better energy efficiency and technology scaling. At the same time, several studies have postulated the potential end of DRAM scaling (or at least a significant slowing down) over the next decade, which further increases the likelihood that DRAM will be replaced by NVRAM memories in future systems.

Inventing the Future—Cross-Disciplinary Holistic System Design

We believe that the confluence of all these trends—the march toward exascale computing and its associated challenges, opportunities related to emerging large-scale distributed data-centric workloads, and potential disruptions from emerging advances in technology—offers us a unique opportunity to rethink traditional system design.

We believe that the next decade of innovation will be characterized by a holistic emphasis that cuts across traditional design boundaries—across layers of design from chips to datacenters; across different fields in computer science, including hardware, systems, and applications; and across different engineering disciplines, including computer engineering, mechanical engineering, and environmental engineering.

We envision that in the future, rather than focusing on the design of single computers, we will focus on the design of computing elements. Specifically, future systems will be (1) composed of simple building blocks that are efficiently co-designed across hardware and software and (2) composed together into computing ensembles, as needed and when needed. We refer to these ideas as designing *disaggregated dematerialized system elements* bound together by a *composable ensemble management* layer. In the discussion below, we present three illustrative examples from our recent research demonstrating the potential for dramatic improvements.

Cross-Layer Power Management

In the past few years, interest has surged in enterprise power management. Given the multifaceted nature of the problem, the solutions have correspondingly focused on different dimensions. For example, some studies have focused on average power reduction for lower electricity costs while others have focused on peak power management for lower air conditioning and power-delivery costs.

Previous studies can also be categorized based on (1) their approaches (e.g., local resource management, distributed resource scheduling, virtual machine migration); (2) options for controlling power (e.g., processor voltage scaling, component sleep states, turning systems off); (3) specific levels of implementation—chip, server, cluster, or datacenter level—hardware, software, or firmware; and (4) objectives and constraints of the optimization problem—for example, whether or not we allow performance loss and whether or not we allow occasional violations in power budgets.

In the future, the focus will be on the design of computing elements, rather than on the design of single computers.

In the future, many (or all) of these solutions are likely to be deployed together to improve coverage and increase power savings. Currently, emergent behavior from the collection of individual optimizations may or may not be globally optimal, or even stable, or even correct! A key need, therefore, is for a carefully designed, flexible, extensible coordination framework that minimizes the need for global information exchange and central arbitration.

In this first example, we explain how a collaborative effort between computer scientists, thermo-mechanical engineers, and control engineering experts led to a novel coordination solution. The solution is summarized in Figure 1 and is further elaborated in Raghavendra et al. (2008). Briefly, this design is based on carefully connecting and overloading the abstractions in current implementations to allow individual controllers to learn and react to the effect of other controllers, the same way they would respond to variations in workload demand. This enables formal mathematical analysis of stability and provides flexibility to dynamic changes in the controllers and system environments. A specific coordination architecture for five individual solutions using different techniques and actuators to optimize for different goals at different system levels across hardware and software demonstrates that a cross-layer solution can achieve significant advantages in correctness, stability, and efficiency over existing state of the art.

Although illustrative design has shown the potential of a cross-disciplinary approach to improving power

management for the cloud, many more opportunities have yet to be explored. Specifically, how do we define the communication and co-ordination interfaces to enable federated architectures? How do we extend solutions to adapt to application-level semantics and heterogeneity in the systems space (Kansal et al., 2009)? How do we design federation at the larger scale typical of cloud systems? Finally, although our discussions have focused on power management, the "intersecting control loops" problem is representative of a larger class of management problems—how architectures generalize to broader resource management domains.

Dematerialized Datacenters

Our second example is a collaborative project by computer scientists, environmental engineers, and mechanical engineers to build a sustainability-aware new datacenter solution. Unlike prior studies that focused purely on operational energy consumption as a proxy for sustainability, we use the metric of *life-cycle exergy destruction* to systematically study the environmental impact of current designs for the entire life cycle of the system, including embedded impact factors related to materials and manufacturing.

A detailed description of exergy is beyond the scope of this article, but briefly, exergy corresponds to the available energy in a system. Unlike energy, which is neither created nor destroyed (the first law of thermodynamics), exergy is continuously consumed in the performance of useful work by any real entropy-generating process (the second law of thermodynamics). Previous studies have shown that the destruction (or consumption) of exergy is representative of the irreversibility associated with various processes. Consequently, at a first-level of approximation, exergy can be used as a proxy to study environmental sustainability.

Studying exergy-efficient designs leads to several new insights (Chang et al., 2010). First, focusing on the most efficient system design does not always produce the most sustainable solution. For example, although energy-proportional designs are optimal in terms of operational electricity consumption, virtual machine

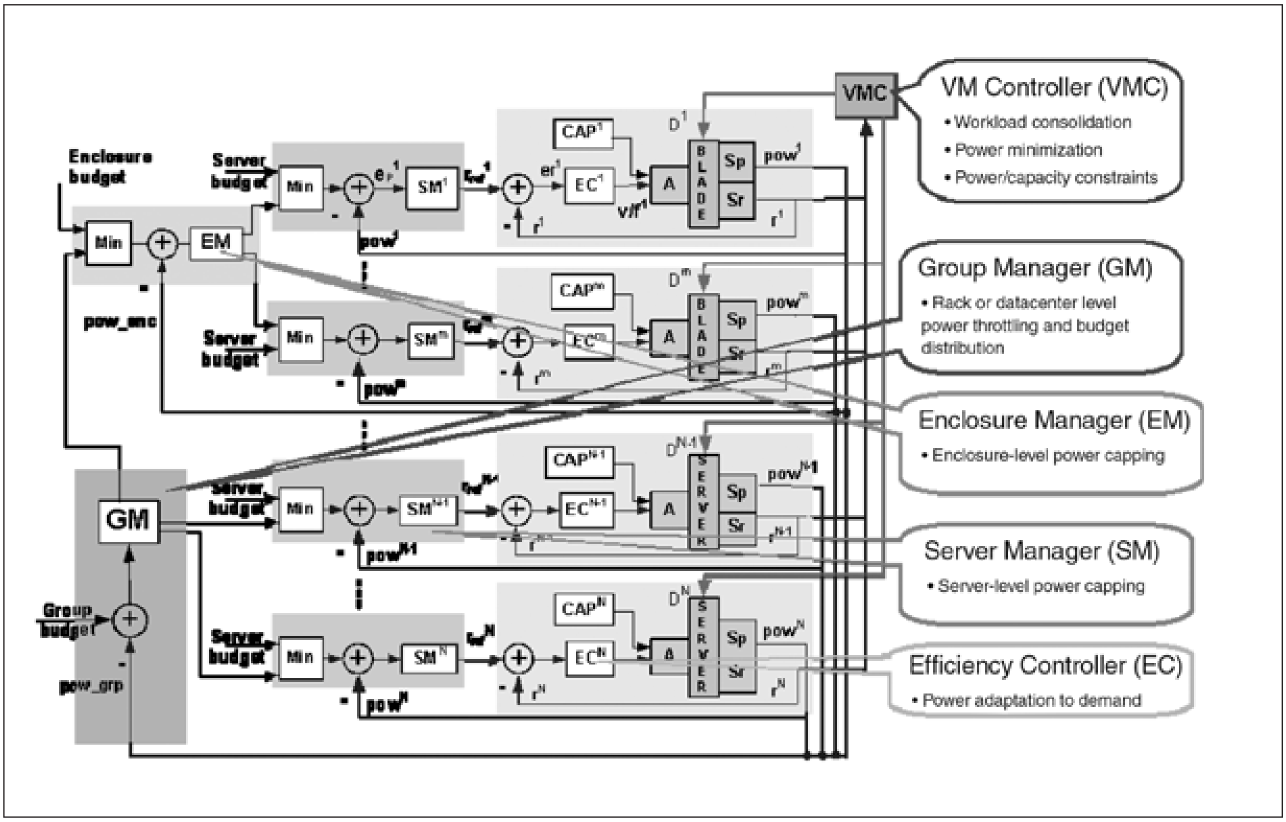


FIGURE 1 A coordinated power-management architecture. The proposed architecture coordinates different kinds of power-management solutions (multiple levels, approaches, time constants, objective functions, and actuators). Key features include (a) a control-theoretic core to enable formal guarantees of stability and (b) intelligent overloading of control channels to include the impact of other controllers, reduce the number of interfaces, and limit the need to access global data.

consolidation is more sustainable than energy proportionality in some cases. Next, the ratio of embedded exergy to total exergy has been steadily increasing over the years, motivating new optimizations that explicitly target embedded exergy (e.g., recycling or dematerialization). Finally, performance and embedded, operational, and infrastructure exergy are not independent variables. Sustainability must be addressed holistically to include them all.

Based on insights provided by the study just described, we propose a new solution (Figure 2) that is co-designed across system architecture and physical organization/packaging. This solution includes three advances that work together to improve sustainability: (1) new material-efficient physical organization, (2) environmentally efficient cooling infrastructures, and (3) effective design of system architectures to enable the reuse of components.

A detailed evaluation of our proposed solution, which includes a combination of sustainability models, computational fluid-dynamics modeling, and full-system computer architecture simulation, demonstrates significant improvements in sustainability, even compared to an aggressive future configuration (Meza et al., 2010). The proposed design illustrates the opportunities that lie ahead. New silicon-efficient architectures, system designs that explicitly target up-cycling, and datacenters with renewable energy sources are the subjects of ongoing research that can bring us closer to truly sustainable datacenters (Patel, 2008).

From Microprocessors to Nanostores

The third example is a cross-disciplinary collaboration among device physicists, computer engineers, and systems software developers to design a disruptive new system architecture for future data-centric workloads (Figure 3). Leveraging the memory-like and disk-like

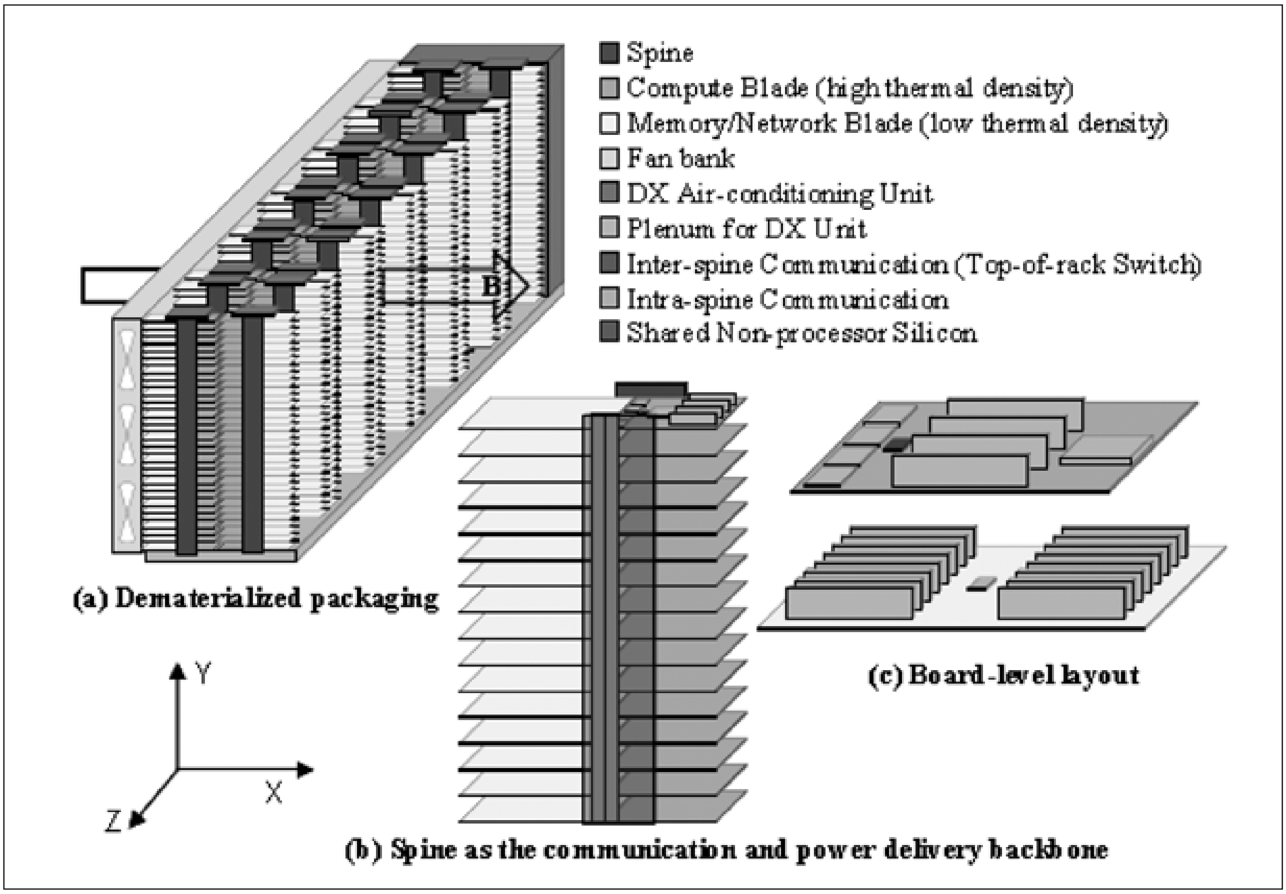


FIGURE 2 Conceptual sketch of a design for a dematerialized green datacenter. This specific design illustrates a container for cloud workloads that incorporates several optimizations co-designed with each other, including (1) new material-efficient physical design, (2) component reuse enabled by a disaggregated system architecture, (3) sharing among collections of systems to reduce the amount of material used in building the system, (4) environmentally efficient cooling that leverages ambient air, and (5) thermal density clustering for lower cooling exergy.

attributes of emerging non-volatile technologies, we propose a new building block, called a nanostore, for data-centric system design (Ranganathan, in press).

A nanostore is a single-chip computer that includes 3-D stacked layers of dense silicon non-volatile memory with a layer of compute cores and a network interface. A large number of individual nanostores can communicate over a simple interconnect and run a data-parallel execution environment like MapReduce to support large-scale distributed data-centric workloads. The two most important aspects of nanostores are (1) the co-location of power-efficient computing with a single-level data store, and (2) support for large-scale distributed design. Together, they provide several benefits.

The single-level data store enables improved performance by providing faster data access (in latency and bandwidth). Energy efficiency is also improved by the flattening of the memory hierarchy and the increased energy efficiency of NVRAM over disk and DRAM. The large-scale distributed design, which increases parallelism and overall data/network bandwidth, allows for higher performance. This design also improves energy efficiency by partitioning the system into smaller elements that can leverage more power-efficient components (e.g., simpler cores).

Our results show that nanostores can achieve orders of magnitude higher performance with dramatically better energy efficiency. More important, they have the potential to be used in new architectural models (e.g., leveraging a hierarchy of computes [Ranganathan, in press]) and to enable new data-centric applications that were previously not possible. Research opportunities include in-systems software optimizations for single-level data stores, new endurance optimizations to improve data reliability, and architectural balance among compute, communication, and storage.

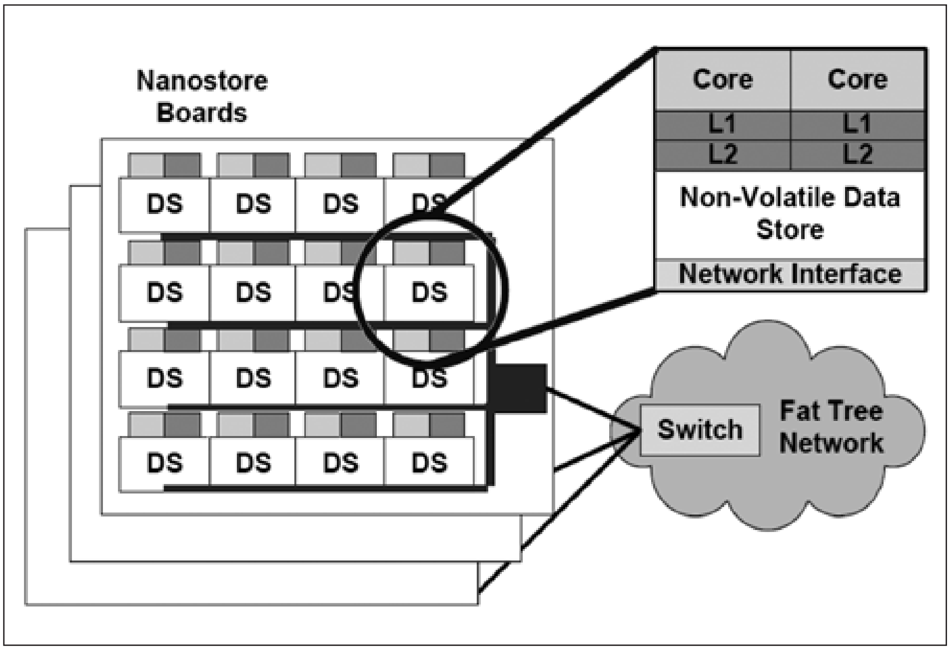


FIGURE 3 The combination of emerging data-centric workloads and upcoming non-volatile and other technologies offer the potential for a new architecture design—"nanostores" that co-locate power-efficient compute cores with non-volatile storage in the same package in a flatter memory hierarchy.

Closing

Although the research described in these examples shows promising results, we believe we have only scratched the surface of what is possible. Opportunities abound for further optimizations, including for hardware-software co-design (e.g., new interfaces and management of persistent data stores [Condit et al., 2009]) and other radical changes in system designs (e.g., bio-inspired "brain" computing [Snider, 2008]).

Overall, the future of computing systems offers rich opportunities for more innovation by the engineering community, particularly for cross-disciplinary research that goes beyond traditional design boundaries. Significant improvements in the computing fabric enabled by these innovations will also provide a foundation for innovations in other disciplines.

References

Banerjee, P., C.D. Patel, C. Bash, and P. Ranganathan. 2009. Sustainable data centers: enabled by supply and demand side management. Design Automation Conference, 2009: 884–887.

Barroso, L.A., and U. Hölzle. 2007. The case for energy proportional computing. IEEE Computer 40(12): 33–37.

Belady, C., A. Rawson, J. Pefleuger, and T. Cader. 2008. Green Grid Data Center Power Efficiency Metrics: PUE

and DCIE. Green Grid white paper #6. Available online at *www.greengrid.org*.

Chang, J., J. Meza, P. Ranganathan, C. Bash, and A. Shah. 2010. Green Server Design: Beyond Operational Energy to Sustainability. Paper presented at HotPower 2010, Vancouver, British Columbia, October 3, 2010.

Condit, J., E.B. Nightingale, C. Frost, E. Ipek, D. Burger, B. Lee, and D. Coetzee. 2009. Better I/O Through Byte-Addressable, Persistent Memory. Presented at Symposium on Operating Systems Principles (SOSP '09), Association for Computing Machinery Inc., October 2009.

Jouppi, N., and Y. Xie. 2009. Emerging technologies and their impact on system design. Tutorial at the International Symposium on Low Power Electronics and Design, 2009.

Kansal, A., J. Liu, A. Singh, R. Nathuji, and T. Abdelzaher. 2009. Semantic-less Coordination of Power Management and Application Performance. In Hotpower 2009 (co-located with SOSP 2009), USENIX, October 10, 2009.

Mayer, M. 2009. The Physics of Data. Talk given at Xerox PARC, August 13, 2009. Available online at *http://www.parc.com/event/936/innovation-at-google.html*.

Meza, J., R. Shih, A. Shah, P. Ranganathan, J. Chang, and C. Bash. 2010. Lifecycle-Based Data Center Design. Presented at ASME International Mechanical Engineering Congress & Exposition, Vancouver, British Columbia, November 14–17, 2010.

Patel, C. 2008. Sustainable IT Ecosystem. Keynote address at the 6th USENIX Conference on File and Storage Technologies, February 26–29, 2008, San Jose, California.

Raghavendra, R., P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu. 2008. No "Power" Struggles: Coordinated Multi-level Power Management for the Data Center. In Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), Seattle, Wash., March 2008.

Ranganathan, P. In press. From microprocessor to nanostores: rethinking system building blocks for the data-centric era. IEEE Computer.

Snider, G., 2008. Memristors as Synapses in a Neural Computing Architecture. Presented at Memristor and Memristive Systems Symposium, University of California, Berkeley, November 21, 2008.

Winter. R. 2008. Why are data warehouses growing so fast? An Update on the Drivers of Data Warehouse Growth. Available online at *http://www.b-eye-network.com/view/7188*.

Research at the intersection of engineering and musical performance includes music cognition, expression synthesis, ensemble coordination, and improvisation.

Demystifying* Music and Its Performance

Elaine Chew

Elaine Chew is associate professor in the Andrew and Erna Viterbi School of Engineering and in the Flora L. Thornton School of Music, University of Southern California. Photo by Brian Morri.

The mathematical nature of music and its imminently quantifiable attributes make it an ideal medium for studying human creativity and cognition. The architecture of musical structures reveals principles of invention and design. The dynamics of musical ensemble offer models of human collaboration. The demands of musical interaction challenge existing computing paradigms and inspire new ones.

Engineering methodology is integral to systematic study, computational modeling, and a scientific understanding of music perception and cognition, as well as to music making. Conversely, understanding the in-action thinking and problem solving integral to music making and cognition can lead to insights into the mechanics of engineering discovery and design. Engineering-music research can also advance commercial interests in Internet radio, music recommendation and discovery, and video games.

The projects described in this article, which originated at the Music Computation and Cognition Laboratory at the University of Southern California, will give readers a sense of the richness and scope of research at the intersection of engineering and musical performance. The research includes computational music cognition, expression synthesis, ensemble coordination, and musical improvisation.

---

* The title is inspired by George Bugliarello's description of "science and engineering as great untanglers of myths" (Bugliarello, 2003).

Analyzing and Visualizing Tonal Structures in Real Time

Most of the music we hear is tonal music, that is, tones (or pitches) organized in time that generate the perception of different levels of stability. The most stable pitch in a musical segment is known as the tonal center, and this pitch gives the key its name. Computational modeling of key-finding dates back to the early days of artificial intelligence (Longuet-Higgins and Steedman, 1971). A popular method for key-finding, devised by Krumhansl and Schmuckler (described in Krumhansl, 1990), is based on the computation of correlation coefficients between the duration profile (vector) of a query stream and experimentally derived probe tone profiles.

Theoretically Efficient Algorithms

In 2000, the author proposed a spiral array model for tonality consisting of a series of nested helices representing pitch classes and major and minor chords and keys. Representations are generated by successive aggregations of their component parts (Chew, 2000). Previous (and most ongoing) research in tonality models has focused only on geometric models (representation) or on computational algorithms that use only the most rudimentary representations. The spiral array attempts to do both. Thus, although the spiral array has many correspondences with earlier models, it can also be applied to the design of efficient algorithms for automated tonal analysis, as well as to the scientific visualization of these algorithms and musical structures (Chew, 2008).

Any stream of notes can generate a center of effect (i.e., a center of gravity of the notes) in the spiral array space. The center of effect generator (CEG) key-finding algorithm based on the spiral array determines the key by searching for the key representation nearest the center of effect of the query stream. The interior-point approach of the CEG algorithm makes it possible to recognize the key quickly and provides a framework for designing further algorithms for automated music analysis.

A natural extension of key-finding is the search for key (or contextual) boundaries. Two algorithms have been proposed for finding key boundaries using the spiral array—one that minimizes the distance between the center of effect of each segment and its closest key (Chew, 2002) and one that finds statistically significant maxima in the distance between the centers of effect of consecutive music segments, without regard to key (Chew, 2005).

The inverse problem of pitch-spelling (turning note numbers, or frequency values, into letter names for music manuscripts that can be read by humans) is essential to automated music transcription. Several variants of a pitch-spelling algorithm using the spiral array have been proposed, such as a cumulative window (Chew and Chen, 2003a), a sliding window (Chew and Chen, 2003b), and a multi-window bootstrapping method (Chew and Chen, 2005).

Robust Working Systems

Converting theoretically efficient algorithms into robust working systems that can stand up to the rigors of musical performance presents many challenges. Using the Software Architecture for Immersipresence framework (François, in press), many of the algorithms described above have been incorporated into the Music on the Spiral Array Real-Time (MuSA.RT) system—an interactive tonal analysis and visualization software (Chew and François, 2003).

MuSA.RT has been used to analyze and visualize music by Pachelbel, Bach, and Barber (Chew and François, 2005). These visualizations have been presented in juxtaposition to Sapp's keyspaces (2005) and Toiviainen's self-organizing maps (2005). MuSA.RT has also been demonstrated internationally and was presented at the AAAS Ig Nobel session in 2008.

Figure 1 shows MuSA.RT in concert at the 2008 ISMIR conference at Drexel University in Philadelphia. As the author plays the music (in this case, Ivan Tcherepnin's *Fêtes–Variations on Happy Birthday)* on the piano, MuSA.RT performs real-time analysis of the tonal patterns. At the same time, visualizations of the tonal structures and trajectories are projected on a large screen, revealing the evolution of tonal patterns—away

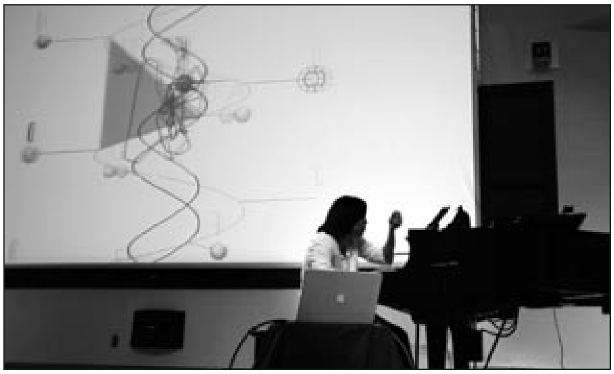


FIGURE 1 MuSA.RT in concert at the 2008 ISMIR conference at Drexel University. Source: © The Philadelphia Enquirer. Reprinted with permission.

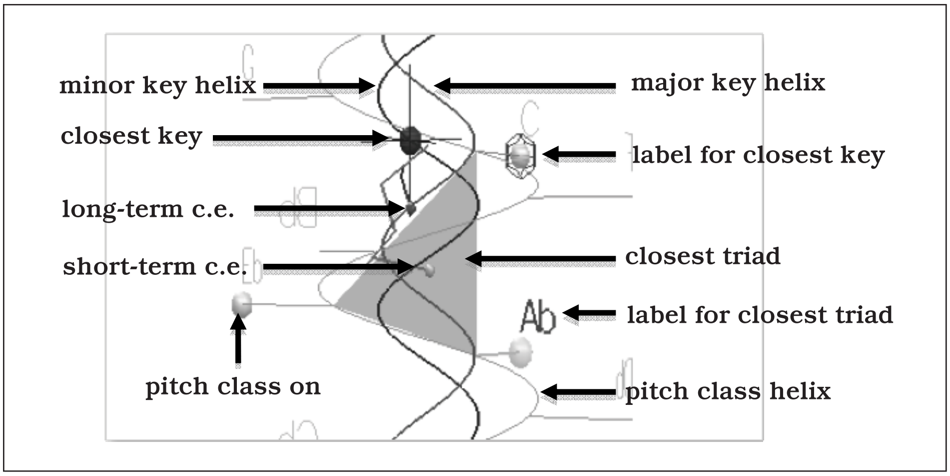


FIGURE 2 Components of the spiral array in MuSA.RT labeled.

from C major and back—over a period of more than ten minutes.

In MuSA.RT version 2.7, the version shown in Figure 2 (as well as Figure 1), the pitch-class helix and pitch names are shown in silver, the major/minor triad helices are hidden, and the major/minor key helices, which are shown in red and blue, respectively, appear as a double helix in the structure's core. When a note is sounded, silver spheres appear on the note name, a short-term center of effect tracks the local context, and the triad closest to it lights up as a pink/blue triangle. A long-term center of effect tracks the larger scale context, and a sphere appears on the closest key, the size of which is inversely proportional to the center of effect-key distance. Lighter violet and darker indigo trails trace the history of the short-term and long-term centers of effect, respectively

An analysis of humor in the music of P.D.Q. Bach (a.k.a. Peter Schickele) by Huron (2004) revealed that many of Schickele's humorous effects are achieved by violating expectations. Using MuSA.RT to analyze P.D.Q. Bach's *Short-Tempered Clavier*, we visualized excessive repetition, as well as incongruous styles, improbable harmonies, and surprising tonal shifts (all of which appeared as far-flung trajectories in the spiral array space) (Chew and François, 2009). Figure 3 shows visualizations of a few of these techniques.

By using the sustain pedal judiciously, and by accenting different notes through duration or stress, a performer can influence the listener's perception of tonal structures. The latest versions of MuSA.RT take into account pedal effects in the computing of tonal

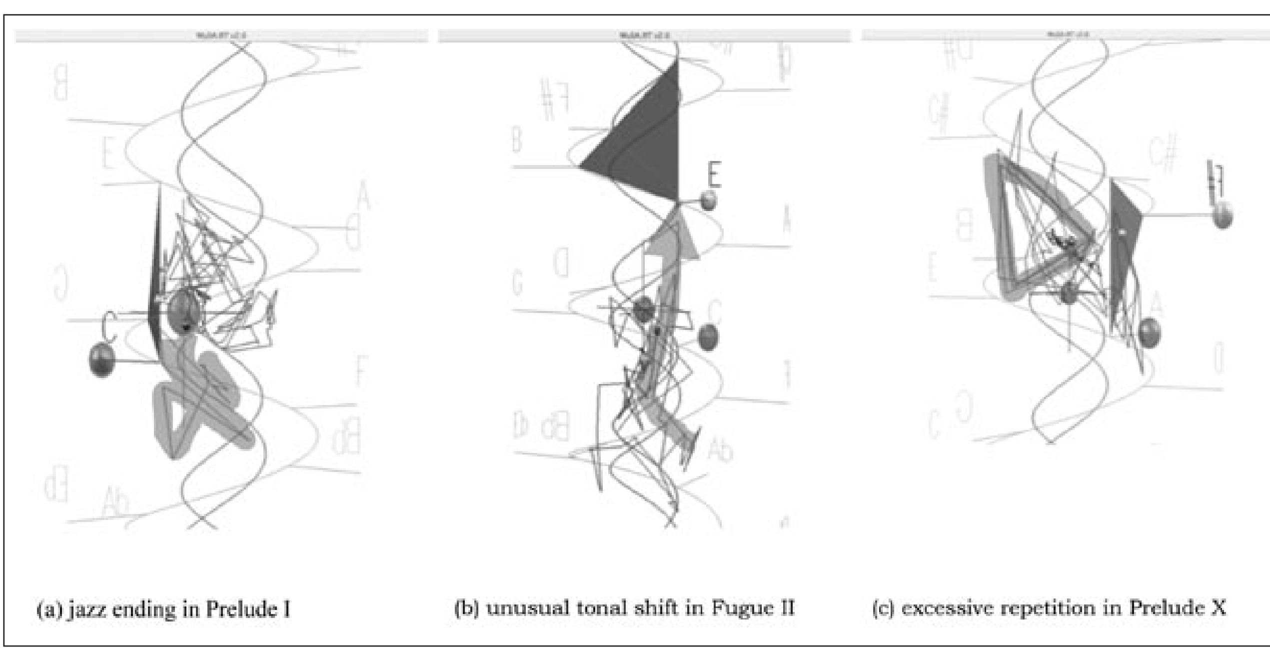


FIGURE 3 Humor devices in P.D.Q. Bach's Short-Tempered Clavier visualized in MuSA.RT. Source: Adapted from Chew and François, 2009.

FIGURE 4 ESP at the USC Festival 125 Pavilion. Photo by Elaine Chew.

structures (Chew and François, 2008). Because center of effect trails react directly to the timing of note soundings, no two human performances result in exactly the same trajectories.

We are currently working on extending spiral array concepts to a higher dimension to represent tetrachords (Alexander et al., in preparation). The resulting pentahelix has a direct correspondence to the orbifold model (a geometric model for representing chords using a topological space with an orbifold structure) (Callendar et al., 2008; Tymoczko, 2006).

Experiencing Music Performance through Driving

Not everyone can play an instrument well enough to execute expressive interpretations at will, but almost everyone can drive a car, at least in a simulation. The Expression Synthesis Project (ESP), based on a literal interpretation of music as locomotion, creates a driving interface for expressive performance that enables even novices to experience the kind of embodied cognition characteristic of expert musical performance (Figure 4).

In ESP (Chew et al., 2005a), the driver uses an accelerator and brake pedal to increase or decrease the speed of the car (music playback). The center line segments in the road approach at one per beat, thus providing a sense of tempo (beat rate and car velocity); this is shown on the speedometer in beats per minute. Suggestions to slow down or speed up are embedded in bends in the road and straight sections, respectively. Thus the road map, which corresponds to an interpretation of a musical piece, often reveals the underlying structure of the piece.

Despite the embedded suggestions, the user is free to choose her/his desired tempo trajectory. In addition, more than one road map (interpretation) can correspond to the same piece of music (Chew et al., 2006). As part of the system design (using SAI), a virtual radius mapping strategy ensures smooth tempo transitions (Figure 5), a hallmark of expert performance, even if the user's driving behavior is erratic (Liu et al., 2006).

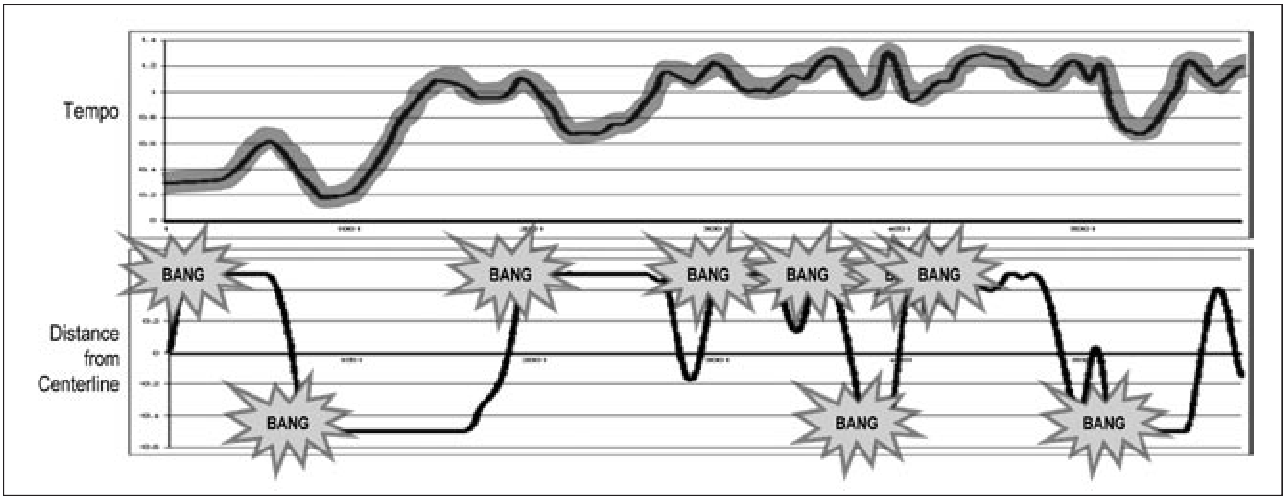


FIGURE 5 Virtual radius mapping ensures smooth tempo transitions despite erratic driving. Source: Adapted from Liu et al., 2006.

FIGURE 6 Delaying each player's sound to his/her own ears to align with incoming audio of the other player's sound.

Charting the Dynamics of Ensemble Coordination

Remote collaboration is integral to our increasingly global and distributed workplaces and society. Musical ensemble performance offers a unique framework through which to examine the dynamics of close collaboration and the challenges of human interaction in the face of network delays.

In a series of distributed immersive performance (DIP) experiments, we recorded the Tosheff Piano Duo performing Poulenc's *Piano Sonata for Four Hands* with auditory delays ranging from 0 milliseconds (ms) to 150 ms. In one experiment, performers heard themselves immediately but heard the other performer with delays. Both performers reported that delays of more than 50 ms caused them to struggle to keep time and compromised their interpretation of the music (Chew et al., 2004).

By delaying each performer's playing to his/her own ears so that it aligned with the incoming signal of the partner's playing (see Figure 6), we created a more satisfying experience for the musicians that allowed them to adapt to the conditions of delay and increased the delay threshold to 65 ms (Chew et al., 2005b), even for the *Final*, a fast and rhythmically demanding movement. Quantitative analysis showed a marked increase in the range of segmental tempo strategies between 50 ms and 75 ms and a marked decline at 100 ms and 150 ms (Chew et al., 2005b).

Most other experiments have treated network delay as a feature of free improvisation rather than a constraint of classical performance. A clapping experiment at Stanford showed that, as auditory delays increased, pairs of musicians slowed down over time. They sped up modestly when delays were less than 11.5 ms (Chafe and Gurevich, 2004). In similar experiments at the University of Rochester, Bartlette et al. (2006) found that latencies of more than 100 ms profoundly impacted the ability of musicians to play as a duo. Using more recent tools and techniques for music alignment and performance analysis, we can now conduct further experiments with the DIP files to create detailed maps of ensemble dynamics (Wolf and Chew, 2010).

On-the-Fly "Architecting" of a Musical Improvisation

Multimodal interaction for musical improvisation (MIMI) was created as a stand-alone performer-centric tool for human-machine improvisation (François et al., 2007). Figure 7 shows MIMI at her concert debut earlier this year.

MIMI takes user input, creates a factor oracle, and traverses it stochastically to generate recombinations of the original input (Assayag and Dubnov, 2004). In previous improvisation systems (Assayag et al., 2006; Pachet, 2003), performers reacted to machine output without prior warning. MIMI allows for more natural interaction by providing visual cues to the origins of the music being generated and by giving musicians a 10-second heads up and review of the musical material.

MIMI's interface allows the performer to decide when the machine learns and when the learning stops, to determine the recombination rate (the degree of fragmenting of the original material), to decide when the machine starts and stops improvising, the loudness of

FIGURE 7 MIMI's concert debut at the People Inside Electronics Concert in Pasadena, California, with Isaac Schankler. Photo by Elaine Chew.

playback, and when to clear the memory (François et al., 2010). By tracking these decisions as the performance unfolds, we can build a record of how an improviser "architects" an improvisation. As work with MIMI continues and performance decisions are documented, our understanding of musical (and hence, human) creativity and design will improve.

Open Courseware

Reviews of music-engineering research as open courseware can be found at *www-scf.usc.edu/~ise575* (Chew, 2006). Each website includes a reading list, presentations and student reviews of papers, and links to final projects. For the 2010 topic in Musical Prosody and Interpretation, highlights of student projects include Brian Highfill's time warping of a MIDI (musical instrument digital interface) file of *Wouldn't It Be Nice* to align with the Beach Boys' recording of the same piece; Chandrasekhar Rajagopal's comparison of guitar and piano performances of Granados' *Spanish Dance No. 5*; and Balamurali Ramasamy Govindaraju's charting of the evolution of vibrato in violin performance over time.

Acknowledgments

This work was supported in part by the National Science Foundation (NSF) under Grants 0321377 and 0347988 and Cooperative Agreement EEC-9529152. Opinions, findings, and conclusions or recommendations expressed in this article are those of the author and not necessarily of NSF.

References

Alexander, S., E. Chew, R. Rowe, and S. Rodriguez. In preparation. The Pentahelix: A Four-Dimensional Realization of the Spiral Array.

Assayag, G., and S. Dubnov. 2004. Using factor oracles for machine improvisation. Soft Computing 8(9): 604–610.

Assayag, G., G. Bloch, M. Chemillier, A. Cont, and S. Dubnov. 2006. Omax Brothers: A Dynamic Topology of Agents for Improvization Learning. In Proceedings of the Workshop on Audio and Music Computing for Multimedia, Santa Barbara, California.

Bartlette, C., D. Headlam, M. Bocko, and G. Velikic. 2006. Effect of network latency on interactive musical performance. Music Perception 24(1): 49–62.

Bugliarello, G. 2003. A new trivium and quadrivium. Bulletin of Science, Technology and Society 23(2): 106–113.

Callendar, C., I. Quinn, and D. Tymoczko. 2008. Generalized voice-leading spaces. Science 320(5874): 346–348.

Chafe, C., and M. Gurevich. 2004. Network Time Delay and Ensemble Accuracy: Effects of Latency, Asymmetry. In Proceedings of the 117th Conference of the Audio Engineering Society, San Francisco, California.

Chew, E. 2000. Towards a Mathematical Model of Tonality. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2000.

Chew, E. 2002. The Spiral Array: An Algorithm for Determining Key Boundaries. Pp. 18–31 in Music and Artificial Intelligence, edited by C. Anagnostopoulou, M. Ferrand, and A. Smaill. Springer LNCS/LNAI, 2445.

Chew, E. 2005. Regards on two regards by Messiaen: post-tonal music segmentation using pitch context distances in the spiral array. Journal of New Music Research 34(4): 341–354.

Chew, E. 2006. Imparting Knowledge and Skills at the Forefront of Interdisciplinary Research—A Case Study on Course Design at the Intersection of Music and Engineering. In Proceedings of the 36th Annual ASEE/IEEE Frontiers in Education Conference, San Diego, California, October 28–31.

Chew, E. 2008. Out of the Grid and into the Spiral: Geometric Interpretations of and Comparisons with the Spiral-Array Model. W. B. Hewlett, E. Selfridge-Field, and E. Correia Jr., eds. Computing in Musicology 15: 51–72.

Chew, E., and Y.-C. Chen. 2003a. Mapping MIDI to the Spiral Array: Disambiguating Pitch Spellings. Pp. 259–275 in Proceedings of the 8th INFORMS Computing Society Conference, Chandler, Arizona, January 8–10, 2003.

Chew, E., and Y.-C. Chen. 2003b. Determining Context-Defining Windows: Pitch Spelling Using the Spiral Array. In Proceedings of the 4th International Conference for Music Information Retrieval, Baltimore, Maryland, October 26–30, 2003.

Chew, E., and Y.-C. Chen. 2005. Real-time pitch spelling using the spiral array. Computer Music Journal 29(2): 61–76.

Chew, E., and A.R.J. François. 2003. MuSA.RT—Music on the Spiral Array. Real-Time. Pp. 448–449 in Proceedings of the ACM Multimedia '03 Conference, Berkeley, California, November 2–8, 2003.

Chew, E., and A.R.J. François. 2005. Interactive multi-scale visualizations of tonal evolution in MuSA.RT Opus 2. ACM Computers in Entertainment 3(4): 16 pages.

Chew, E. and A.R.J. François. 2008. MuSA.RT and the Pedal: the Role of the Sustain Pedal in Clarifying Tonal Structure. In Proceedings of the Tenth International Conference on Music Perception and Cognition, Sapporo, Japan, August 25–29, 2008.

Chew, E., and A.R.J. François. 2009. Visible Humour—Seeing P.D.Q. Bach's Musical Humour Devices in *The Short-Tempered Clavier* on the Spiral Array Space. In Mathematics and Computation in Music—First International Conference, edited by T. Noll and T. Klouche. Springer CCIS 37.

Chew, E., A.R.J. François, J. Liu, and A. Yang. 2005a. ESP: A Driving Interface for Musical Expression Synthesis. In Proceedings of the Conference on New Interfaces for Musical Expression, Vancouver, B.C., Canada, May 26–28, 2005.

Chew, E., R. Zimmermann, A. Sawchuk, C. Papadopoulos, C. Kyriakakis, C. Tanoue, D. Desai, M. Pawar, R. Sinha, and W. Meyer. 2005b. A Second Report on the User Experiments in the Distributed Immersive Performance Project. In Proceedings of the 5th Open Workshop of MUSICNETWORK, Vienna, Austria, July 4–5, 2005.

Chew, E., J. Liu, and A.R.J. François. 2006. ESP: Roadmaps as Constructed Interpretations and Guides to Expressive Performance. Pp. 137–145 in Proceedings of the First Workshop on Audio and Music Computing for Multimedia, Santa Barbara, California, October 27, 2006.

Chew, E., R. Zimmermann, A.A. Sawchuk, C. Kyriakakis, C. Papadopoulos, A.R.J. François, G. Kim, A. Rizzo, and A. Volk. 2004. Musical Interaction at a Distance: Distributed Immersive Performance. In Proceedings of the 4th Open Workshop of MUSICNETWORK, Barcelona, Spain, September 15–16, 2004.

Chew, E., A. Sawchuk, C. Tanoue, and R. Zimmermann. 2006. Segmental Tempo Analysis of Performances in Performer-Centered experiments in the Distributed Immersive Performance Project. In Proceedings of International Conference on Sound and Music Computing, Salerno, Italy, November 24–26, 2005.

François, A.R.J. In press. An architectural framework for the design, analysis and implementation of interactive systems. The Computer Journal.

François, A.R.J., E. Chew, and C.D. Thurmond. 2007. Visual Feedback in Performer-Machine Interaction for Musical Improvisation. In Proceedings of International Conference on New Instruments for Musical Expression, New York, June 2007.

François, A.R.J., I. Schankler, and E. Chew. 2010. MIMI4x: An Interactive Audio-Visual Installation for High-Level Structural Improvisation. In Proceedings of the Workshop on Interactive Multimedia Installations and Digital Art, Singapore, July 23, 2010.

Huron, D. 2004. Music-engendered laughter: an analysis of humor devices in P.D.Q. Bach. Pp. 700–704 in Proceedings of the International Conference on Music Perception and Cognition, Evanston, Ill.

Krumhansl, C. 1990. Cognitive Foundations of Musical Pitch. Oxford, U.K.: Oxford University Press.

Liu, J., E. Chew, and A.R.J. François. 2006. From Driving to Expressive Music Performance: Ensuring Tempo Smoothness. In Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, Hollywood, California, June 14–16. 2006.

Longuet-Higgins, H.C., and M. Steedman. 1971. On interpreting Bach. Machine Intelligence 6: 221–241.

Pachet, F. 2003. The continuator: musical interaction with style. Journal of New Music Research 32(3): 333–341.

Sapp, C. 2005. Visual hierarchical key analysis. ACM Computers in Entertainment 3(4): 19 pages.

Toiviainen, P. 2005. Visualization of tonal content with self-organizing maps and self-similarity matrices. ACM Computers in Entertainment 3(4): 10 pages.

Tymoczko, D. 2006. The geometry of musical chords. Science 313(5783): 72–74.

Wolf, K., and E. Chew. 2010. Evaluation of Performance-to-Score MIDI Alignment of Piano Duets. In online abstracts of the International Conference on Music Information Retrieval, Utrecht, The Netherlands, August 9–13, 2010.

Aerospace scientists tend not to trust autonomous software for projects on which many years and large sums of money have been spent.

Intelligent Autonomy in Robotic Systems

Mark Campbell is associate professor, Sibley School of Mechanical and Aerospace Engineering, Cornell University.

Mark Campbell

Automation is now apparent in many aspects of our lives, from aerospace systems (e.g., autopilots) to manufacturing processes (e.g., assembly lines) to robotic vacuum cleaners. However, although many aerospace systems exhibit some autonomy, it can be argued that such systems could be far more advanced than they are. For example, although autonomy in deep-space missions is impressive, it is still well behind autonomous ground systems. Reasons for this gap range from proximity to the hardware and environmental hardships to scientists tending not to trust autonomous software for projects on which many years and dollars have been spent.

Looking back, the adoption of the autopilot, an example of advanced autonomy for complex systems, aroused similar resistance. Although autopilot systems are required for most commercial flights today, it took many years for pilots to accept a computer flying an aircraft.

Factors that influence the adoption of autonomous systems include reliability, trust, training, and knowledge of failure modes. These factors are amplified in aerospace systems where the environment/proximity can be challenging and high costs and human lives are at stake. On deep-space missions, for example, which take many years to plan and develop and where system failure can often be traced back to small failures, there has been significant resistance to the adoption of autonomous systems.

In this article, I describe two improvements that can encourage more robust autonomy on aerospace missions: (1) a deeper level of intelligence

in robotic systems; and (2) more efficient integration of autonomous systems and humans.

Intelligence in Robotics

Current robotic systems work very well for repeated tasks (e.g., in manufacturing). However, their long-term reliability for more complex tasks (e.g., driving a car) is much less assured. Interestingly, humans provide an intuitive benchmark, because they perform at a level of deep intelligence that typically enables many complex tasks to be performed well. However, this level of intelligence is difficult to emulate in software. Characteristics of this deep intelligence include learning over time, reasoning about and overcoming uncertainties/new situations as they arise, and developing long-term strategies.

Researchers in many areas are investigating the concept of deeper intelligence in robotics. For our purposes, we look into three research topics motivated in part by tasks that humans perform intelligently: (1) tightly integrated perception, anticipation, and planning; (2) learning; and (3) verified plans in the presence of uncertainties.

Tightly Integrated Perception, Anticipation, and Planning

As robotic systems have matured, an important advancement has been the development of high-throughput sensors. Consider, for example, Cornell's autonomous driving vehicle, one of only six that completed the 2007 DARPA Urban Challenge (DUC) (Figure 1). The vehicle (robot) has a perception system with a 64-scan lidar unit (100Mbits/sec), 4-scan lidar units (10MBits/sec), radars, and cameras (1,200Mbits/sec).

Although the vehicle's performance in the DUC was considered a success (Iagnemma et al., 2008; Miller et al., 2008), there were many close calls, several small collisions, and a number of human-assisted restarts. In fact, the fragility of practical robotic intelligence was apparent when many simple mistakes in perception cascaded into larger failures.

One critical problem was the mismatch between perception, which is typically *probabilistic* because sensors yield data that are inherently uncertain compared to the true system, and planning, which is *deterministic* because plans must be implemented in the real world. To date, perception research typically provides robotic planners with probabilistic "snapshots" of the environment, which leads to "reactive," rather than "intelligent," behaviors in autonomous robots.

Aerospace systems have similar problems. Figure 2 shows a cooperative unmanned air vehicle (UAV) system for searching out and tracking objects of interest (Campbell and Whitacre, 2007), such as tuna fish or survivors of hurricanes and fires. System failures include searching only part of an area, losing track of objects when they move out of sight (e.g., behind a tree or under a bridge), or vibrations or sensor uncertainty aboard the aircraft.

Overcoming these problems will require new theory that provides tighter linkage between sensors/probabilistic perception and actions/planning (Thrun et al., 2005). Given the high data throughput of the sensors on most systems, a key first step is to convert "data to information." This will require fusing data from many sensors to provide an accurate picture of the static, but potentially dynamic, environment, including terrain type and the identity and behaviors of obstacles (Diebel and Thrun, 2006; Schoenberg et al., 2010).

Take driving, for example. A human is very good at prioritizing relatively small amounts of information

FIGURE 1 The autonomous driving vehicle developed by Cornell University, which can handle large amounts of data intelligently. Left: Robot with multi-modal sensors for perceiving the environment. Middle: Screenshot of a 64-scan lidar unit. Right: Screenshot of a color camera.

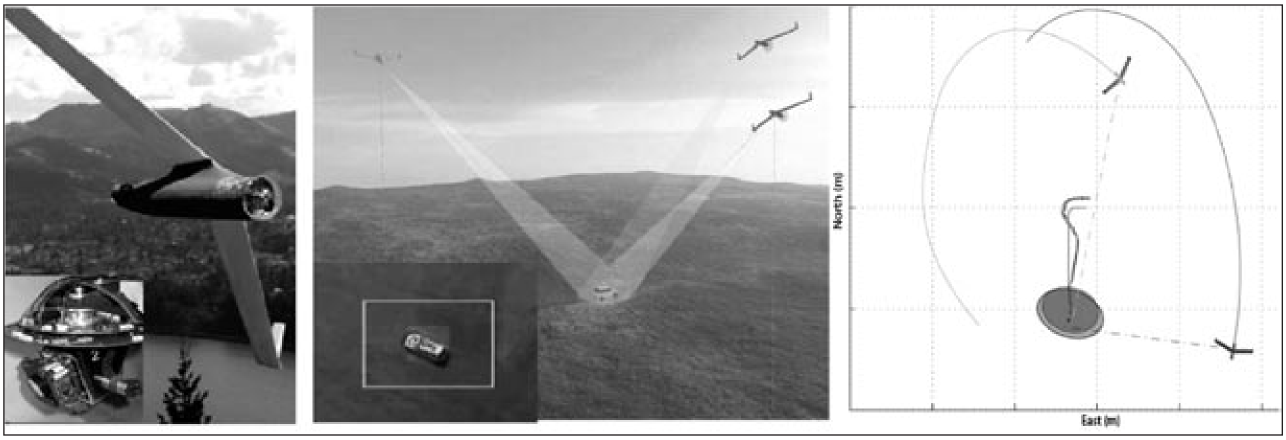


FIGURE 2 Multiple UAV system. Left: SeaScan UAV with a camera-based turret. Center: Notional illustration of cooperative tracking using UAVs. Right: Flight test data of two UAVs tracking a truck over a communication network with losses (e.g., dropped packets).

(i.e., from the eyes), as well as a priori learned models. If an object is far away, the human typically focuses on the "gist" of the scene, such as object type (Ross and Oliva, 2010). If an object is closer, such as when something is about to hit a car, the primary focus is on proximity, rather than type (Cutting, 2003). To ensure that large amounts of data are transformed into critical information that can be used in decision making, we need new representations and perception methods, particularly methods that are computationally tractable.

Plans must then be developed based on probabilistic information. Thus, the second essential step is to convert "information to decisions," which will require a new paradigm to ensure that planning occurs to a particular level of probability, while also incorporating changes in the environment, such as the appearance of objects (from the perceived information and a priori models). This is especially important in dynamic environments, where the behavior and motion of objects are strongly related to object type.

For autonomous driving (Figure 1), important factors for planning include the motion of other cars, cyclists, and pedestrians in the context of a map (Blackmore et al., 2010; Hardy and Campbell, 2010; Havlak and Campbell, 2010). For cooperative UAVs (Figure 2), important factors include the motion of objects and other UAVs (Grocholsky et al., 2004; Ousingsawat and Campbell, 2007). Although humans typically handle these issues well by relying on learned models of objects, including their motions and behaviors, developing robotic systems that can handle these variables reliably can be computationally demanding (McClelland and Campbell, 2010).

For single and cooperative UAV systems, such as those used for search and rescue or defense missions, data are typically in the form of optical/infra red video and lidar. The necessary information includes detecting humans, locating survivors in clutter, and tracking moving cars—even if there are visual obstructions, such as trees or buildings. Actions based on this information then include deciding where to fly, a decision strongly influenced by sensing and coverage, and deciding what information to share (among UAVs and/or with ground operators). Ongoing work in sensor fusion and optimization-based planning have focused on these problems, particularly as the number of UAVs increases (Campbell and Whitacre, 2007; Ousingsawat and Campbell, 2007).

Learning

Humans typically drive very well because they learn *safely* over time (rules, object types and motion, relative speeds, etc.). However, for robots, driving well is very challenging, especially when uncertainties are prevalent. Consider Figure 3, which shows a map of the DUC course, with an overlay of 53 instances of emergency slamming of brakes by Cornell's autonomous vehicle. Interestingly, many of these braking events occurred during multiple passes near the same areas; the most frequent (18 times) took place near a single concrete barrier jutting out from the others, making it appear (to the perception algorithms) that it was another car (Miller et al., 2008).

Currently, a number of researchers exploring learning methods (e.g., Abbeel et al., 2010) have developed algorithms that learn helicopter dynamics/maneuver

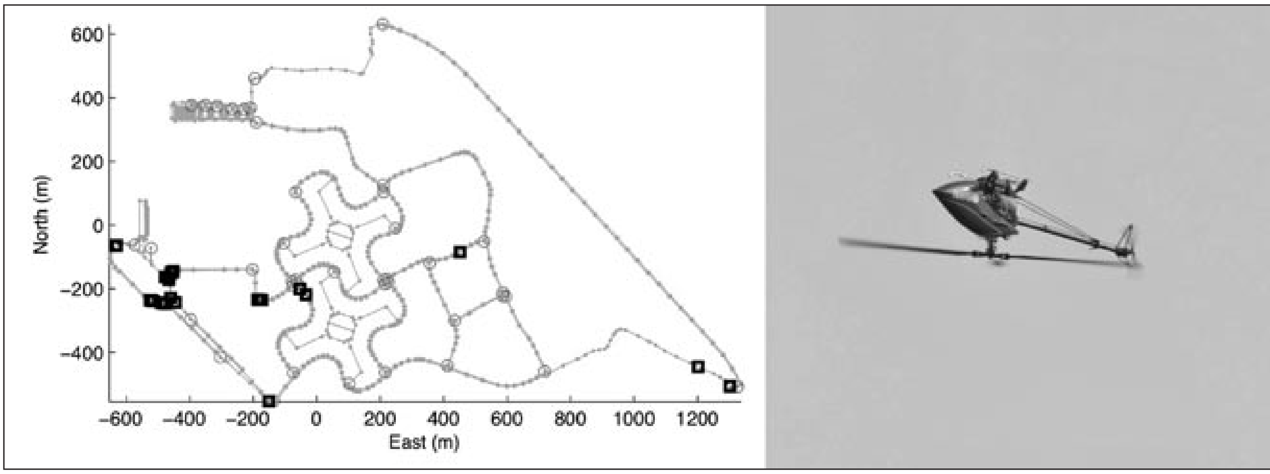


FIGURE 3 Left: Map of the DUC course (lines = map; circles = stop signs). Black squares indicate where brakes were applied quickly during the six-hour mission. Right: A model helicopter (operated by remote control or as an autonomous vehicle) in mid-maneuver. These complex maneuvers can be learned from an expert or by experimentation. Photo by E. Fratkin.

models over time from data provided by an expert pilot (Figure 3). Although learning seems straightforward to humans, it is difficult to implement algorithmically. New algorithms must be developed to ensure safe learning over time and adjust to new environments or uncertainties that have not been seen before (e.g., if at some point a car really did appear).

Verification and Validation in the Presence of Uncertainties

Current methods of validating software for autonomy in aerospace systems involve a series of expensive evaluation steps to heuristically develop confidence in the system. For example, UAV flight software typically requires validation first on a software simulator, then on a hardware-in-the-loop simulator, and then on flight tests. Fault-management systems continue to operate during flights, as required.

Researchers in formal logic, model checkers, and control theory have recently developed a set of tools that capture specification of tasks using more intuitive language/algorithms (Kress-Gazit et al., 2009; Wongpiromsarn et al., 2009). Consider, for example, a car driving through an intersection with another car in the area (Figure 4). The rules of the road can be specified by logic, and controllers for autonomous driving can be automatically generated. Current research, however, typically addresses only simple models with little or no uncertainty.

New theory and methods will be necessary to incorporate uncertainties in perception, motion, and actions into a verifiable planning framework. Logic specifications must provide *probabilistic* guarantees of the high-level behavior of the robot, such as provably safe autonomous driving 99.9 percent of the time. These methods are also obviously important for aerospace systems, such as commercial airplanes and deep-space missions, where high costs and many lives are at risk.

Interaction between Humans and Robots

Although interaction between humans and robots is of immense importance, it typically has a soft theoretical background. Human-robotic interaction, as it is typically called, includes a wide range of research. For example, tasks must be coordinated to take advantage of the strengths of both humans and robots; theory must scale well with larger teams; humans must not become overloaded or bored; and external influences, such as deciding if UAVs will have the ability to make actionable decisions or planetary rovers will be able to make scientific decisions, must be taken into consideration.

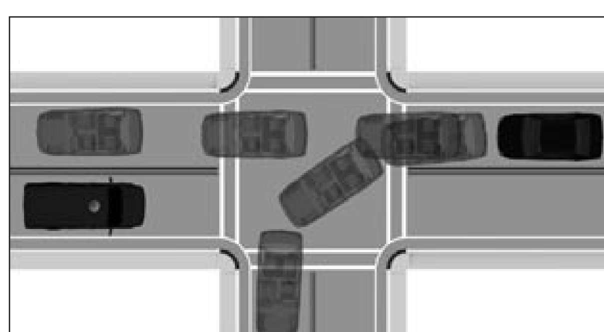


FIGURE 4 Example of using probabilistic anticipation for provably safe plans in autonomous driving.

Efficient integration of autonomy with humans is essential for advanced aerospace systems. For example, a robot vacuuming a floor requires minimal interaction with humans, but search and tracking using a team of UAVs equipped with sensors and weapons is much more challenging, not only because of the complexity of the tasks and system, but also because of the inherent stress of the situation.

The subject of interactions between humans and robots has many aspects and complexities. We look at three research topics, all of which may lead to more efficient and natural integration: (1) fusion of human and robotic information; (2) natural, robust, and high-performing interaction; and (3) scalable theory that enables easy adoption as well as formal analysis.

Fusion of Human and Robotic Information

Humans typically provide high-level commands to autonomous robots, but clearly they can also contribute important information, such as an opinion about which area of Mars to explore or whether a far off object in a cluttered environment is a person or a tree. Critical research is being conducted using machine-learning methods to formally model human opinions/decisions as sources of uncertain information and then fuse it with other information, such as information provided by the robot (Ahmed and Campbell, 2008, 2010).

Figure 5 shows a search experiment with five humans, each of whom has a satellite map overlaid with a density function that probabilistically captures the "location" of objects (Bourgault et al., 2008). The human sensor, in this case, is relatively simple: yes/no detection. A model of the human sensing process was developed by having the humans locate objects at various locations relative to their positions and look vectors. Intuitively, the ability to detect an object declines with increasing range and peripheral vision.

The fusion process, however, must also include uncertain measurements of the human's location and look vector. During the experiment, each human moved to a separate area, while fusing his/her own (uncertain) sensory information to create an updated density function for the location of objects. Fusion with information from other humans occurred when communication allowed—in this case, only at close range. Figure 5 shows the trajectory of one human's path and the real-time fused density of object location.

This experiment demonstrated initial decision modeling and fusion results, but the human decision was decidedly simple. To be useful, however, research, particularly in the area of machine learning, must model more complex outputs, such as strategic decisions over time or decisions made with little a priori data. New methods will be necessary to fuse information from many different sources, such as a human classifying items based on a discrete set of objects or pointing to a continuous area, or combinations of the two.

Natural, Robust, High-Performing Interaction

For effective teamwork by humans and robots, it is important to understand the strengths and weaknesses of both. Humans can provide critical strategic analyses but are subject to stress and fatigue, as well as boredom. In addition, they may have biases that must be taken into account (Parasuraman et al., 2000; Shah et al., 2009). Robots can perform repetitious tasks

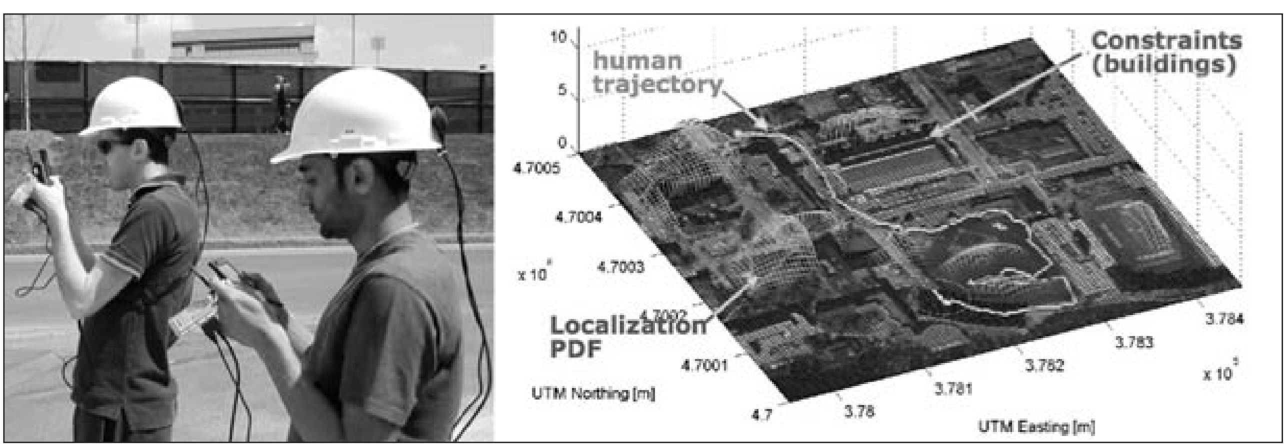


FIGURE 5 Search experiment with a network of five humans. Left: Humans with hand-held PCs, local network, GPS, and compass. Right: Overlay of satellite imagery with a density of "probable" locations.

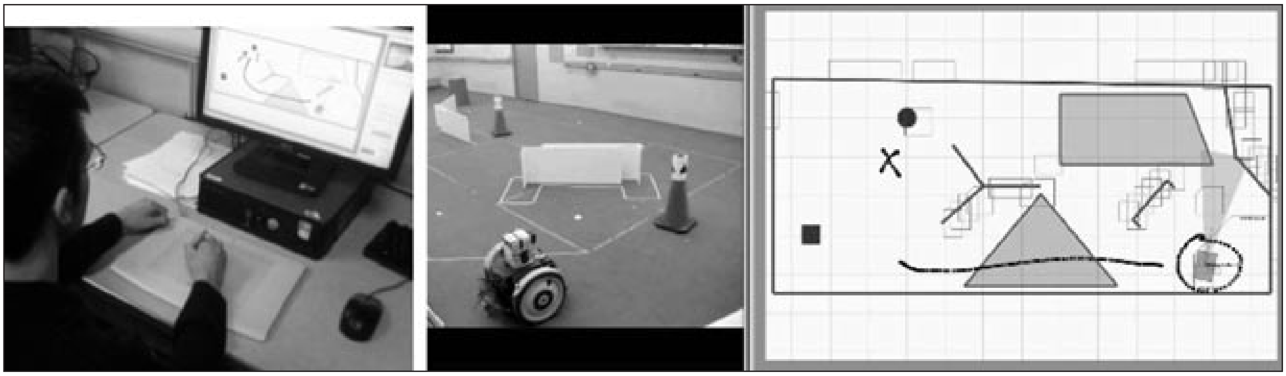


FIGURE 6 Human-drawn gesture commands for robots. Left: Human operator at a tablet-based computer. Center: A robot exploring an environment. Right: A screen shot showing how the human selected a robot, drew a potential path, and selected an area to explore.

without bias or feelings. The strengths and weaknesses of both are, in effect, constraints in the design of systems in which humans and robots must work together seamlessly.

The most common interaction is through a computer, such as a mouse, keyboard, and screen. Sometimes, however, a human operator may be presented with more critical information for monitoring or decision making than he/she can handle. For example, the operator may be monitoring two UAVs during a search mission, and both may require command inputs at the same time. Or a human operator who becomes bored when monitoring video screens for hours at a time may not respond as quickly or effectively as necessary when action is required.

Taking advantage of recent commercial developments in computers that allow humans to interact with systems in many ways, current research is focused on multi-modal interaction. For example, Finomore et al. (2007) explored voice and chat inputs. Shah and Campbell (2010) are focusing on drawing commands on a tablet PC (Figure 6), where pixels are used to infer the "most probable" commands. The human operator can override a command if it is not correct, and the next most probable command will be suggested. Results of this study have shown a high statistical accuracy in recognizing the correct command by the human (Shah and Campbell, 2010).

More advanced systems are also being developed. Kress-Gazit and colleagues (2008) have developed a natural language parser that selects the appropriate command from spoken language and develops a provably correct controller for a robot. Boussemart and Cummings (2008) and Hoffman and Breazeal (2010) are working on modeling the human as a simplified, event-based decision maker; the robot then "anticipates" what the human wants to do and makes decisions appropriately. Although the latter approach is currently being applied only to simplified systems, it has the potential to improve team performance. Even non-traditional interfaces, such as commanding a UAV by brain waves, are being investigated (Akce et al., 2010).

Scalable Theory

A key constraint on the development of theory and implementations of teams of humans and robots is being able to scale up the theory to apply to large numbers (McLoughlin and Campbell, 2007; Sukkarieh et al., 2003). This is particularly important in defense applications, where hundreds, sometimes thousands of humans/vehicles must share information and plan together.

Most of the focus has been on hierarchical structures, but fully decentralized structures might also be effective (Ponda et al., 2010). Recent research has focused almost exclusively on large teams of cooperative vehicles, but given some level of human modeling, these methods could work for human/robot teams as well. The testing and adoption of these approaches, which will necessarily depend partly on cost and reliability, will continue to be challenging.

References

Abbeel, P., A. Coates, and A.Y. Ng. 2010. Autonomous helicopter aerobatics through apprenticeship learning. International Journal of Robotics Research 29(7). doi: 10.1177/0278364910371999.

Ahmed, N., and M. Campbell. 2008. Multimodal Operator Decision Models. Pp. 4504–4509 in American Control Conference 2008. New York: IEEE.

Ahmed, N., and M. Campbell. 2010. Variational Bayesian Data Fusion of Multi-class Discrete Observations with Applications to Cooperative Human-Robot Estimation. Pp. 186–191 in 2010 IEEE International Conference on Robotics and Automation. New York: IEEE.

Akce, A., M. Johnson, and T. Bretl. 2010. Remote Teleoperation of an Unmanned Aircraft with a Brain-Machine Interace: Theory and Preliminary Results. Presentation at 2010 International Robotics and Automation Conference. Available online at *http://www.akce.name/presentations/ICRA2010.pdf*.

Blackmore, L., M. Ono, A. Bektassov, and B.C. Williams. 2010. A probabilistic particle-control approximation of chance-constrained stochastic predictive control. IEEE Transactions on Robotics, 26(3): 502–517.

Bourgault, F., A. Chokshi, J. Wang, D. Shah, J. Schoenberg, R. Iyer, F. Cedano, and M. Campbell. 2008. Scalable Bayesian Human-Robot Cooperation in Mobile Sensor Networks. Pp. 2342–2349 in IEEE/RS International Conference on Intelligent Robots and Systems. New York: IEEE.

Boussemart, Y., and M.L. Cummings. 2008. Behavioral Recognition and Prediction of an Operator Supervising Multiple Heterogeneous Unmanned Vehicles. Presented at Humans Operating Unmanned Systems, Brest, France, September 3–4, 2008. Available online at *http://web.mit.edu/aeroastro/labs/halab/papers/boussemart%20-%20humous08-v1.3-final.pdf*.

Campbell, M.E., and W.W. Whitacre. 2007. Cooperative tracking using vision measurements on SeaScan UAVs. IEEE Transactions on Control Systems Technology 15(4): 613–626.

Cutting, J.E. 2003. Reconceiving Perceptual Space. Pp. 215–238 in Perceiving Pictures: An Interdisciplinary Approach to Pictorial Space, edited by H. Hecht, R. Schwartz, and M. Atherton. Cambridge, Mass.: MIT Press.

Diebel, J., and S. Thrun. 2006. An application of Markov random fields to range sensing. Advances in Neural Information Processing Systems 18: 291.

Finomore, V.S., B.A. Knott, W.T. Nelson, S.M. Galster, and R.S. Bolia. 2007. The Effects of Multimodal Collaboration Technology on Subjective Workload Profiles of Tactical Air Battle Management Teams. Technical report #AFRL-HE-WP-TP-2007-0012. Wright Patterson Air Force Base, Defense Technical Information Center.

Grocholsky, B., A. Makarenko, and H. Durrant-Whyte. 2004. Information-Theoretic Coordinated Control of Multiple Sensor Platforms. Pp. 1521–1526 in Proceedings of the IEEE International Conference on Robotics and Automation, Vol. 1. New York: IEEE.

Hardy, J., and M. Campbell. 2010. Contingency Planning over Hybrid Obstacle Predictions for Autonomous Road Vehicles. Presented at IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, October 18–22, 2010.

Havlak, F., and M. Campbell. 2010. Discrete and Continuous, Probabilistic Anticipation for Autonomous Robots in Urban Environments. Presented at SPIE Europe Conference on Unmanned/Unattended Sensors and Sensor Networks, Toulouse, France, September 20–23, 2010.

Hoffman, G., and C, Breazeal. 2010. Effects of anticipatory perceptual simulation on practiced human-robot tasks. Autonomous Robots 28: 403–423.

Iagnemma, K., M. Buehler, and S. Singh, eds. 2008. Special Issue on the 2007 DARPA Urban Challenge, Part 1. Journal of Field Robotics 25(8): 423–860.

Kress-Gazit, H., G.E. Fainekos, and G.J. Pappas. 2008. Translating structured English to robot controllers. Advanced Robotics: Special Issue on Selected Papers from IROS 2007 22(12): 1343–1359.

Kress-Gazit, H., G.E. Fainekos, and G.J. Pappas. 2009. Temporal logic-based reactive mission and motion planning. IEEE Transactions on Robotics 25(6): 1370–1381.

McClelland, M., and M. Campbell. 2010. Anticipation as a Method for Overcoming Time Delay in Control of Remote Systems. Presented at the AIAA Guidance, Navigation and Control Conference, Toronto, Ontario, Canada, August 2–5, 2010. Reston, Va.: AIAA.

McLoughlin, T., and M. Campbell. 2007. Scalable GNC architecture and sensor scheduling for large spacecraft networks. AIAA Journal of Guidance, Control, and Dynamics 30(2): 289–300.

Miller, I., M. Campbell, D. Huttenlocher, A. Nathan, F.-R. Kline, P. Moran, N. Zych, B. Schimpf, S. Lupashin, E. Garcia, J. Catlin, M. Kurdziel, and H. Fujishima. 2008. Team Cornell's Skynet: robust perception and planning in an urban environment. Journal of Field Robotics 25(8): 493–527.

Ousingsawat, J., and M.E. Campbell. 2007. Optimal cooperative reconnaissance using multiple vehicles. AIAA Journal of Guidance Control and Dynamics 30(1): 122–132.

Parasuraman, R., T.B. Sheridan, and C.D. Wickens. 2000. A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 30: 286–297.

Ponda, S., H.L. Choi, and J.P. How. 2010. Predictive Planning for Heterogeneous Human-Robot Teams. Presented at AIAA Infotech@Aerospace Conference 2010, Atlanta, Georgia, April 20–22, 2010.

Ross, M., and A. Oliva. 2010. Estimating perception of scene layout. Journal of Vision 10: 1–25.

Schoenberg, J., A. Nathan, and M. Campbell. 2010. Segmentation of Dense Range Information in Complex Urban Scenes. Presented at IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, October 18–22, 2010.

Shah, D., and M. Campbell. 2010. A Robust Sketch Interface for Natural Robot Control. Presented at IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, October 18–22, 2010.

Shah, D., M. Campbell, F. Bourgault, and N. Ahmed. 2009. An Empirical Study of Human-Robotic Teams with Three Levels of Autonomy. Presented at AIAA Infotech@Aerospace Conference 2009, Seattle, Washington, April 7, 2009.

Sukkarieh, S., E. Nettleton, J.H. Kim, M. Ridley, A. Goktogan, and H. Durrant-Whyte. 2003. The ANSER Project: data fusion across multiple uninhabited air vehicles. The International Journal of Robotics Research 22(7–8): 505.

Thrun, S., W. Burgard, and D. Fox. 2005. Probabilistic Robotics. Cambridge, Mass.: MIT Press.

Wongpiromsarn, T., U. Topcu, and R.M. Murray. 2009. Receding Horizon Temporal Logic Planning for Dynamical Systems. Available online at *http://www.cds.caltech.edu/~utopcu/images/1/10/WTM-cdc09.pdf*.

In the near future, fully autonomous unmanned aircraft systems will be both technologically feasible and safe.

Certifiable Autonomous Flight Management for Unmanned Aircraft Systems

Ella M. Atkins is associate professor, Aerospace Engineering Department, University of Michigan.

Ella M. Atkins

The next-generation air transportation system (NextGen) will achieve unprecedented levels of throughput[1] and safety by judiciously integrating human supervisors with automation aids. NextGen designers have focused their attention mostly on commercial transport operations, and few standards have been proposed for the burgeoning number of unmanned aircraft systems (UAS).[2] In this article, I describe challenges associated with the safe, efficient integration of UAS into the National Airspace System (NAS).

Current Aircraft Automation

Although existing aircraft autopilots can fly from takeoff through landing, perhaps the most serious technological impediment to fully autonomous flight is proving their safety in the presence of anomalies such as unexpected traffic, onboard failures, and conflicting data. Current aircraft automation is "rigid" in that designers have favored simplicity over adaptability. As a result, responding in emergency situations, particularly following events

---

[1] In the context of NextGen, "throughput" is defined as the number of aircraft that can be moved through a particular airspace per unit time.

[2] The aerospace community has adopted the term "unmanned aircraft system" (UAS) to replace "unmanned air vehicle" (UAV), because a contemporary unmanned aircraft is a complex "system" of tightly integrated hardware and software that typically supports data acquisition, processing, and communication rather than cargo transport.

that degrade flight performance (e.g., a jammed control surface, loss of engine thrust, icing, or damage to the aircraft structure) requires the intervention and ingenuity of a human pilot or operator.

If the automation system on a manned aircraft proves to be insufficient, the onboard flight crew is immersed in an environment that facilitates decision making and control. Furthermore, modern aircraft rarely experience emergencies because their safety-critical systems are designed with triple redundancy.

Manned flights are much more costly and have a bigger environmental impact than autonomous flights.

Ensuring Safety in Unmanned Aircraft Systems

To be considered "safe," UAS operations must maintain acceptable levels of risk to other aircraft and to people and property. An unmanned aircraft may actually fly "safely" throughout an accident sequence as long as it poses no risk to people or property on the ground or in the air. Small UAS are often considered expendable in that they do not carry passengers and the equipment itself may have little value. Thus, if a small UAS crashes into unimproved terrain, it poses a negligible risk to people or property.

UAS cannot accomplish the ambitious missions for which they are designed, however, if we limit them to operating over unpopulated regions. To reap the benefits of UAS, we must develop and deploy technologies that decrease the likelihood of a UAS encountering conditions that can lead to an incident or accident.

However, the recipe for safety on manned aircraft is impractical for small UAS. First, triple redundancy for all safety-critical systems would impose unacceptable cost, weight, and volume constraints for small aircraft. Second, although transport aircraft typically fly direct routes to deliver their payloads, surveillance aircraft are capable of dynamically re-planning their flight trajectories in response to the evolving mission or to observed data (e.g., the detection of a target to be tracked).

Finally, UAS are operated remotely, and operators are never directly engaged in a situation in which their lives are at risk. In fact, operators can only interact with a UAS via datalink, and "lost link" is currently one of the most common problems.[3]

Safety Challenges for Small Unmanned Aircraft

With limited redundancy, highly dynamic routes, and strictly remote supervision, small UAS face formidable automation challenges. As the number of unmanned aircraft increases and as safety-oriented technology development continues to lag behind the development of new platforms, mission capabilities, and operational efficiency (e.g., one operator for multiple vehicles), it is becoming increasingly urgent that these issues be addressed. In addition, a large user base for UAS is emerging, which includes military and homeland security missions and commercial ventures.

Making the routine operation of unmanned aircraft safe wherever they are needed will substantially reduce the need for costlier manned flights that have a much greater adverse impact on the environment. However, for unmanned aircraft to operate near other aircraft or over populated areas, they must be capable of managing system failures, lost links, and dynamic routing, including collision avoidance, in a way that is "safe" for people and property.

We are currently working to augment autonomous decision making in the presence of actuator or sensor failures by expanding the definition of "flight envelope" to account for evolving physical, computational, perceptual, and environmental constraints. The flight envelope is traditionally defined by physical constraints, but under damage or failure conditions the envelope can contract. An autonomous flight controller must be capable of identifying and respecting these constraints to minimize the risk of loss-of-control as the aircraft continues on its mission or executes a safe emergency landing.

The autonomous flight manager can minimize risk by following flight plans that maximize safety margins first and then maximize traditional efficiency metrics (e.g., energy or fuel use). Thus flight plans for UAS may first divert the aircraft away from populated regions on the ground or densely occupied airspace and then decide whether to continue a degraded flight plan or end the mission through intentional flight termination or a

[3] This issue was recently brought to our attention when a Fire Scout UAS aircraft lost its communication link and inappropriately flew quite close to restricted airspace around Washington, D.C. (*http://www.nytimes.com/2010/08/26/us/26drone.html?_r=4&partner=rss&emc=rss*).

controlled landing in a nearby safe (unpopulated) area. The key to certification of this autonomous decision-making will be guaranteeing that acceptable risk levels, both real and perceived, are maintained.

Addressing Safety Challenges

In the discussion that follows, we look first at the problem of certifiable autonomous UAS flights in the context of current flight and air traffic management (ATM) technologies, which are primarily designed to ensure safe air transportation with an onboard flight crew. In this context, we also describe current and anticipated roles for automation and human operators.

Next, we characterize emerging UAS missions that are driving the need for fully autonomous flight management and integration into the NAS. Because loss-of-control is a major concern, I suggest an expanded definition of the flight envelope in the context of a real-life case study, the dual bird strike incident of US Airways Flight 1549 in 2009. That incident highlighted the need for enhanced automation in emergency situations for both manned and unmanned aircraft.

Finally, challenges to certification are summarized and strategies are suggested that will ultimately enable UAS to fly, autonomously, in integrated airspace over populated as well as rural areas.

Flight and Air Traffic Management: A System-of-Systems

In the NextGen NAS, avionics systems onboard aircraft will be comprised of a complex network of processing, sensing, actuation, and communication elements (Atkins, 2010a). UAS, whether autonomous or not, must be certified to fit into this system. All NextGen aircraft will be networked through datalinks to ATM centers responsible for coordinating routes and arrival/departure times.

The Federal Aviation Administration (FAA) and its collaborators have proposed a system-wide information management (SWIM) architecture (*www.swim.gov*) that will enable collaborative, flexible decision-making for all NAS users; it is assumed that all NextGen aircraft will be capable of accurately following planned 4-D trajectories (three-dimensional positions plus times), maintaining separation from other traffic, and sharing pertinent information such as GPS coordinates, traffic alerts, and wind conditions. Protocols for system-wide and aircraft-centric decision-making must be established to handle adverse weather conditions, encounters with wake turbulence, and situations in which other aircraft deviate from their expected routes.

To operate efficiently in controlled NextGen airspace, all aircraft will be equipped with an onboard flight management system (FMS) that replicates current functionality, including precise following of the approved flight plan, system monitoring, communication, and pilot interfaces (Fishbein, 1995; Liden, 1994). Automatic Dependent Surveillance–Broadcast (ADS-B) systems will also communicate aircraft status information (e.g., position, velocity) to ensure collision avoidance. Without such equipment, it will be difficult to guarantee that traffic remains separated throughout flight, especially when manned and unmanned aircraft are involved.

Low-Cost Flight Management Systems

Small operators, from general and sports aviation to unmanned aircraft, will require low-cost options to the current FMS. Although advanced miniaturized electronics can make low-cost, lightweight FMS possible (Beard, 2010), producing and marketing these systems will require a concerted effort in the face of potentially slim profit margins and formidable validation and verification requirements.

> *Small aircraft, manned or autonomous, will require low-cost options to the current flight management system.*

The current FMS can devise and follow a flight plan from origin to destination airport. In the future, automation in both manned and unmanned aircraft is expected to include making and coordinating dynamic routing decisions based on real-time observations (e.g., weather), other traffic, or even mission goals (e.g., target tracking). Quite simply, we are rapidly moving toward collaborative human-machine decision making or fully autonomous decision-making rather than relying on human supervisors of autonomous systems, particularly if operators are not onboard.

From Lost Link to Optional Link

Today's unmanned aircraft are flown by remote pilots/operators who designate waypoints or a sequence of

waypoints, as well as a rendezvous location. However, as was mentioned above, communication (lost link) failure is a common and challenging unresolved issue for UAS. Addressing this problem will require that developers not only improve the availability of links, but simultaneously pursue technological advances that will render links less critical to safety.

As the level of autonomy increases to support extended periods of operation without communication links, UAS must be able to operate "unattended" for extended periods of time, potentially weeks or months, and to collect and disseminate data without supervision unless the mission changes.

Sense-and-Avoid Capability

Because human pilots cannot easily see and avoid smaller UAS, "sense and avoid" has become a top priority for the safe integration of UAS into NAS. A certified sense-and-avoid technology will provide another step toward fully autonomous or unattended flight management.

Emerging Unmanned Missions

A less-studied but critical safety issue for UAS operations as part of NAS is maintaining safe operations in the presence of anomalies. Researchers are beginning to study requirements for autonomously carrying out UAS missions (Weber and Euteneuer, 2010) with the goal of producing automation technology that can be certified safe in both nominal and conceivable off-nominal conditions. In this section, we focus on the "surveillance" missions that distinguish UAS—particularly small unmanned aircraft that must operate at low cost in sparsely populated airspace—from traditional transport operations.

Traditional Transport Operations

Traditional transport aircraft have a single goal—to fly a human or cargo payload safely from an origin to a destination airport with minimal cost to the airline. The "best" routes are, therefore, direct, with vectors around traffic or weather as needed. Schedules can be negotiated up to flight time, and passengers and cargo carriers expect on-time delivery, as costs increase with delay. In the context of autonomous transport UAS (e.g., cargo carriers), issues include loss of facilities or adverse weather at the destination airport, failure or damage conditions (e.g., loss of fuel or power) that render the destination unreachable, and security issues that result in a system-wide change in flight plans (e.g., temporary flight restrictions).

Unmanned Surveillance Aircraft

Unlike traditional transport aircraft, the goal of surveillance unmanned aircraft may be to search a geographical region, to loiter over one or more critical sites, or to follow a surveillance target along an unpredictable route. A summary of potential commercial applications (Figure 1) that complement the myriad of military uses for surveillance flights, shows that surveillance and support are the primary emerging mission categories that will require the expansion of existing NAS protocols to manage dynamic routing and the presence of UAS in (1) uncontrolled, low-altitude airspace currently occupied primarily by general aviation aircraft and (2) congested airport terminal areas where traffic is actively managed (Atkins et al., 2009).

This will mean that UAS will mix with the full fleet of manned operations, ranging from sports and

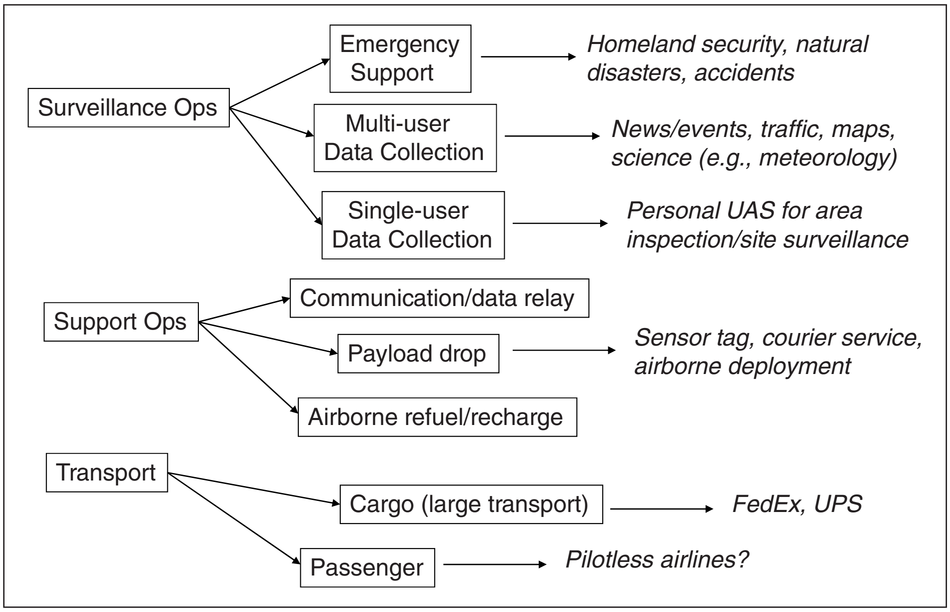


FIGURE 1 Emerging commercial applications for unmanned aircraft. Source: Atkins et al., 2009.

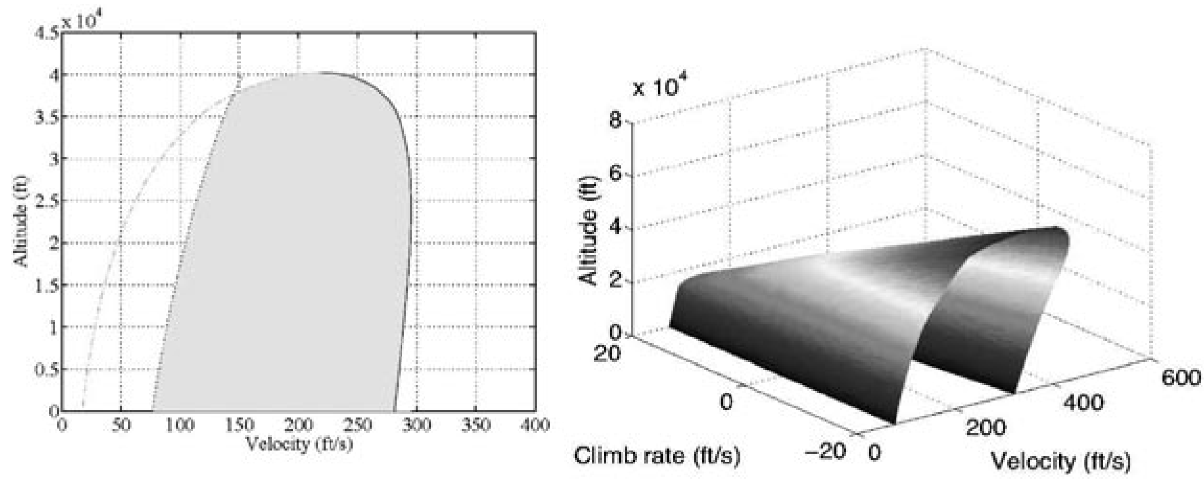


FIGURE 2 Examples of steady-level (left) and 3-D (right) traditional flight envelopes. Source: McClamroch, in press.

recreational aircraft operated by pilots with limited training to jets carrying hundreds of passengers. UAS missions also will overfly populated areas for a variety of purposes, such as monitoring traffic, collecting atmospheric data over urban centers, and inspecting sites of interest. Even small unmanned aircraft have the capacity to provide support for communication, courier services, and so on.[4]

Unmanned aircraft can work in formations that can be modeled and directed as a single entity by air traffic controllers. This capability can give controllers much more leeway in sequencing and separating larger sets of traffic than would be possible if all UAS flights were considered distinct.

UAS teams may also negotiate tasks but fly independent routes, such as when persistent long-term coverage is critical to a successful mission or when cooperative coverage from multiple angles is necessary to ensure that a critical ground target is not lost in an urban environment. Some activities may be scheduled in advance and prioritized through equity considerations (e.g., traffic monitoring), but activities related to homeland security or disaster response are unscheduled and may take priority even over airline operations.

Although the effects of high-altitude UAS must be taken into account by NAS, low-altitude aircraft operating over populated regions or in proximity to major airports will be the most challenging to accommodate in the NextGen NAS. UAS must, of course, be safe, but they must also be fairly accommodated through the extension of NAS metrics (e.g., access, capacity, efficiency, and flexibility) so they can handle operations when persistent surveillance over a region of interest is more important than equitable access to a congested airport runway.

Extending the Flight Envelope to Minimize the Risk of Loss-of-Control

Loss-of-control, the most frequent cause of aviation accidents for all vehicle classes, occurs when an aircraft exits its nominal flight envelope making it impossible to follow its desired flight trajectory (Kwatny et al., 2009). Current autopilot systems rely on intuitive, linearized, steady-flight models (Figure 2) that reveal how aerodynamic stalls and thrusts constrain the flight envelope (McClamroch, in press).

To ensure the safe operation of UAS and to prove that autonomous system performance is reliable, an FMS for autonomous aircraft capable of *provably* avoiding loss-of-control in all situations where avoidance is possible will be essential. This will require that the autonomous system understand its flight envelope sufficiently to ensure that its future path only traverses "stabilizable" flight states (i.e., states the autonomous controller can achieve or maintain without risking loss-of-control).

Researchers are beginning to develop nonlinear system-identification and feedback control algorithms that offer stable, controlled flight some distance beyond the nominal "steady flight" envelope (Tang et al., 2009). Such systems could make it feasible for an

[4] Large cargo carriers would also benefit from flying unmanned aircraft, which require only base personnel who would not have to be located at potential departure sites.

autonomous system to "discover" this more expansive envelope (Choi et al., 2010) and continue stable operation despite anomalies in the environment (e.g., strong winds) or onboard systems (e.g., control-surface failures or structural damage) that would otherwise lead to loss-of-control.

Flight Envelope Discovery

Figure 3 shows the flight envelope for an F-16 with an aileron jammed at 10 degrees. In this case, the aircraft can only maintain steady straight flight at slow speeds. The traversing curve shows an example of a flight envelope discovery process incrementally planned as the envelope is estimated from an initial high-speed turning state through stabilizable states to a final slow-speed straight state (Yi and Atkins, 2010).

This slow-speed, gentle-descent final state and its surrounding neighborhood are appropriate for final approach to landing, indicating that the aircraft can safely fly its approach as long as it remains within the envelope. Once the envelope has been identified, a landing flight plan guaranteed to be feasible under the condition of the control surface jam can be automatically generated.

Figure 4 illustrates the emergency flight management sequence of discovering the degraded flight envelope, selecting a nearby landing site, and constructing a feasible flight plan to that site. Although a runway landing site is presumed in the figure, an off-runway site would probably be selected for a small UAS that required little open space for landing. The sequence in Figure 4 mirrors the emergency procedures a pilot would follow when faced with degraded performance. Note that all of the steps in this process could be implemented autonomously with existing technology.

Autonomous Reaction to Off-Nominal Conditions

The remaining challenge is to prove that such an autonomous system is capable of recognizing and reacting to a sufficient range of off-nominal situations to be considered "safe" without a human pilot as backup. To illustrate how autonomous emergency flight management could improve safety, we investigated the application of our emergency flight planning algorithms to the 2009 Hudson River landing (Figure 5) of US Airways Flight 1549 (Atkins, 2010b).

About two minutes after the aircraft departed from LaGuardia (LGA) Airport in New York, it encountered a flock of large Canada geese. Following multiple bird strikes, the aerodynamic performance of the aircraft was unchanged, but propulsive power was no longer available because of the ingestion of large birds into both jet

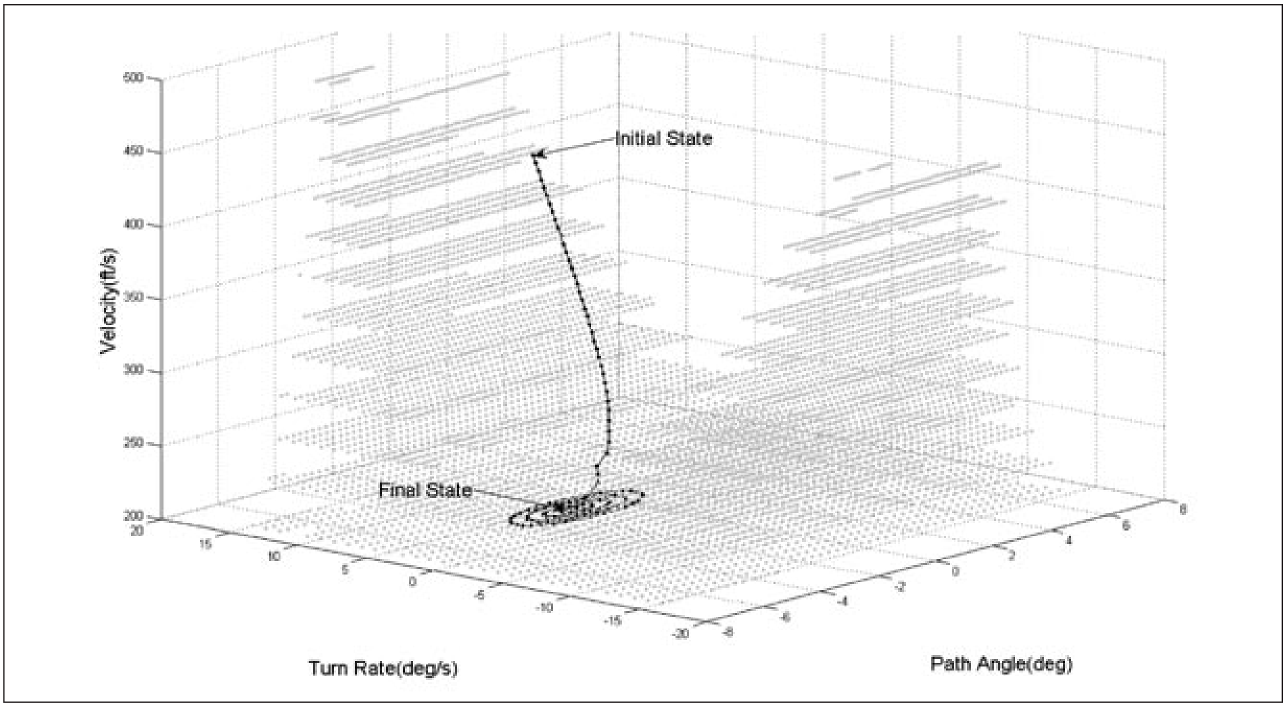


FIGURE 3 Trim-state discovery for an F-16 with a 10-degree aileron jam. Source: Yi and Atkins, 2010.

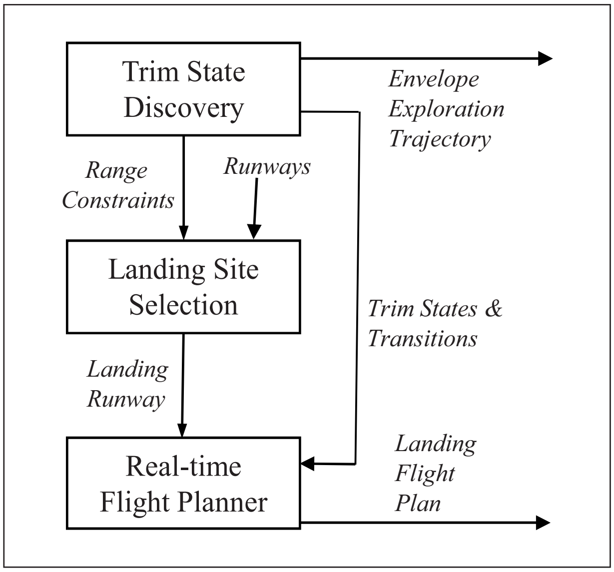


FIGURE 4 Simplified version of an emergency flight-planning sequence for a jet with degraded performance.

engines, which forced the aircraft to glide to a landing. In this event, the pilot aptly glided the plan to a safe landing on the Hudson River. All passengers and crew survived, most with no injuries, and the flight crew has been rightly honored for its exemplary performance.

In the case of Flight 1549, our adaptive flight planner first identified the glide (no-thrust) footprint from the coordinates at which thrust was initially lost. This analysis indicated that the aircraft could return to LGA as long as the return was initiated rapidly, before too much altitude was lost. Our landing site search algorithm prioritized LGA runway 31 as the best choice because of its headwind, but runways 13 and 22 were also initially reachable.

Figure 6 illustrates the feasible landing trajectories

FIGURE 5 Post-landing photo of US Airways Flight 1549 in the Hudson River (*http://www.wired.com/images_blogs/autopia/2010/01/us_airways_1549_cropped.jpg*).

for Flight 1549 automatically generated in less than one second by our pre-existing engine-out flight planner adapted to Airbus A320 glide and turn capabilities. Notably, runway 31 was reachable only if the turn back to LGA was initiated within approximately 10 seconds after the incident. Runways 13 and 22 were reachable for another 10 seconds, indicating that the pilot (or autopilot if available) did in fact have to initiate the return to LGA no more than approximately 20 seconds after thrust was lost.

We believe that if an automation aid had been available to rapidly compute and share the safe glide trajectory back to LGA, and if datalink coordination with air traffic control had been possible to facilitate clearing LGA departure traffic, Flight 1549 could have returned to LGA and avoided the very high-risk (albeit successful in this case) water landing. In short, this simple, provably correct "glide to landing" planning tool represents a substantial, technologically sound improvement over the level of autonomous emergency flight management available today and is a step toward the more ambitious goal of fully autonomous flight management.

Certification of Fully Autonomous Operation

Every year the FAA is asked to certify a wide variety of unmanned aircraft for flight in the NAS. Although most unmanned operations are currently conducted over remote regions where risks to people and property are minimal, certification is and must continue to be based on guarantees of correct responses in nominal conditions, as well as contingency management to ensure safety. Although redundancy will continue to be key to maintaining an acceptable level of risk of damage to people and property in the event of failures, for UAS aircraft, triple redundancy architecture as is present in commercial transport aircraft may not be necessary because ditching the aircraft is often a viable option.

Safety certification is a difficult process that requires some trust in claims by manufacturers and operators about aircraft design and usage. Automation algorithms, however, can ultimately be validated through rigorous mathematical and simulation-based verification processes to provide quantitative measures of robustness, at least for envisioned anomalies in weather, onboard systems, and traffic.

Addressing Rigidity in Flight Management Systems

The remaining vulnerability of a fully autonomous UAS FMS is its potential rigidity, which could lead to

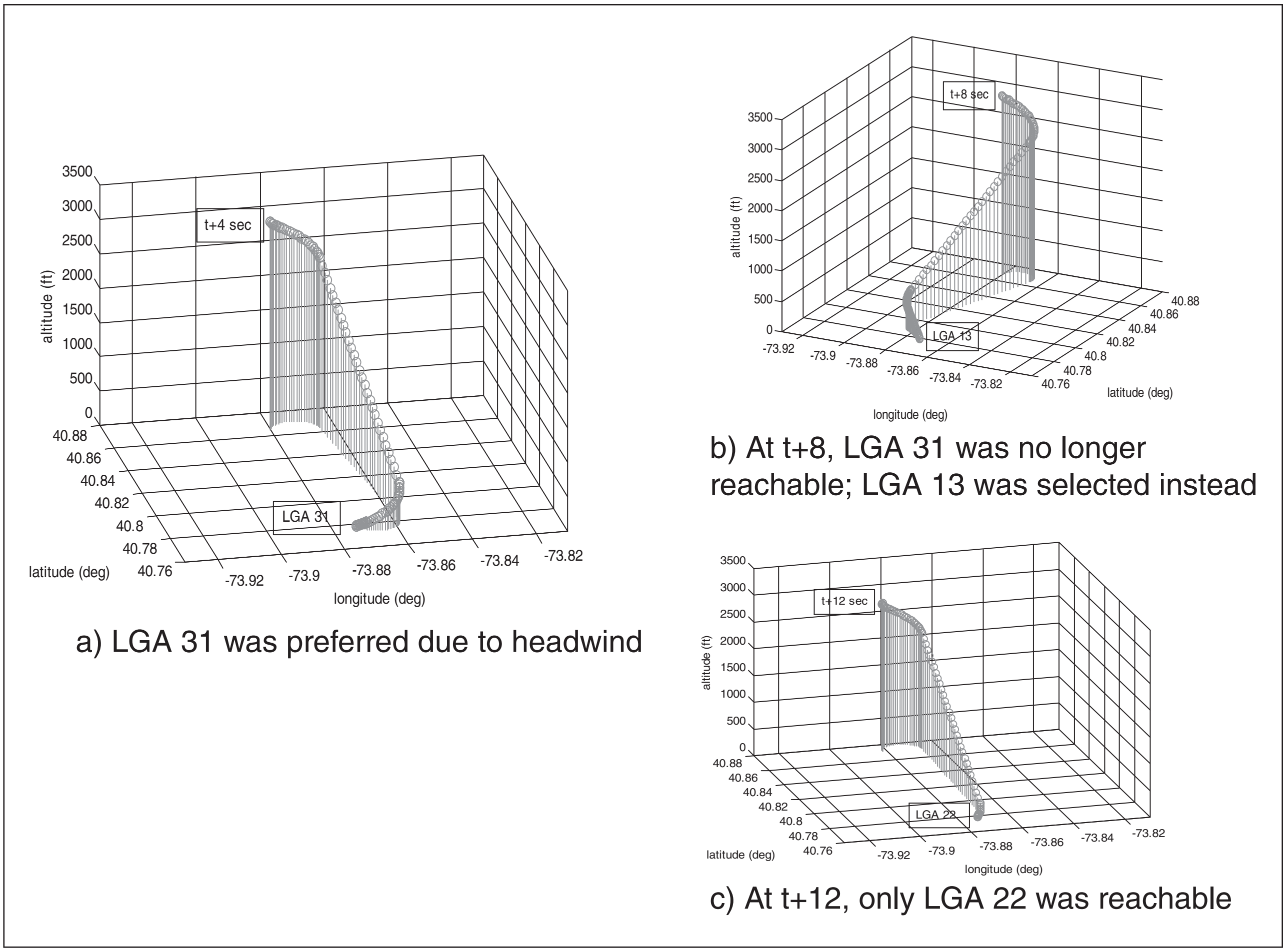


FIGURE 6 Feasible trajectories for Flight 1549 to return to LaGuardia Airport. Source: Atkins, 2010b.

an improper response in a truly unanticipated situation. The default method for managing this vulnerability has been to insert a human pilot into the aircraft control loop. However, with remote operators who have limited engagement with the aircraft, human intervention may not be the best way to offset automation rigidity. If that is the case, the certification of fully autonomous UAS FMS must be based on meeting or exceeding human capabilities, although assessing the human capacity for response will be challenging.

For remote unmanned aircraft, we can start by characterizing the bounds on user commands. Formal methods of validating and verifying automation algorithms and their implementations, as well as assessing their flexibility (rigidity), will also be essential. Simulation and flight testing will, of course, be necessary to gain trust, but we propose that simulation should be secondary to formal proofs of correctness when assessing the performance, robustness, and ultimately safety of autonomous UAS.

Conclusion

Ultimately, fully autonomous UAS operation will be both technologically feasible and safe. The only remaining issue will be overcoming public perceptions and lack of trust, which we believe can be mitigated by long-term exposure to safe and beneficial UAS operations.

References

Atkins, E., A. Khalsa, and M. Groden. 2009. Commercial Low-Altitude UAS Operations in Population Centers. Presented at the American Institute of Aeronautics and Astronautics (AIAA) Aircraft Technology, Integration, and Operations Conference, Hilton Head, South Carolina, September 2009.

Atkins, E. 2010a. Aerospace Avionics Systems. Chapter 391 in Encyclopedia of Aerospace Engineering, edited by R. Blockey and W. Shyy. New York: Wiley

Atkins, E. 2010b. Emergency Landing Automation Aids: An Evaluation Inspired by US Airways Flight 1549. Presented

at the AIAA Infotech@Aerospace Conference, Atlanta, Georgia, April 2010.

Beard, R. 2010. Embedded UAS Autopilot and Sensor Systems. Chapter 392 in Encyclopedia of Aerospace Engineering, edited by R. Blockey and W. Shyy. New York: Wiley.

Choi, H., E. Atkins, and G. Yi. 2010. Flight Envelope Discovery for Damage Resilience with Application to an F16. Presented at the AIAA Infotech@Aerospace Conference, Atlanta, Georgia, April 2010.

Fishbein, S.B. 1995. Flight Management Systems: The Evolution of Avionics and Navigational Technology. Santa Barbara, Calif.: Praeger.

Kwatny, H., J. Dongmo, B. Cheng, G. Bajpai, M. Yawar, and C. Belcastro. 2009. Aircraft Accident Prevention: Loss-of-Control Analysis. In Proceedings of the Guidance, Navigation, and Control Conference, Chicago, Ill., August 2009. Reston, Va.: AIAA.

Liden, S. 1994. The Evolution of Flight Management Systems. Pp. 157–169 in Proceedings of the IEEE Digital Avionics Conference. New York: IEEE.

McClamroch, N.H. In press. Steady Aircraft Flight and Performance.

Tang, L., M. Roemer, J. Ge, A. Crassidis, J. Prasad, and C. Belcastro. 2009. Methodologies for Adaptive Flight Envelope Estimation and Protection. In Proceedings of the Guidance, Navigation, and Control Conference, AIAA, Chicago, Ill., August 2009. Reston, Va.: AIAA.

Weber, R., and E. Euteneuer. 2010. Avionics to enable UAS integration into the NextGen ATS. Presented at AIAA Guidance, Navigation, and Control Conference, Toronto, Ontario, August 2010.

Yi, G., and E. Atkins. 2010. Trim State Discovery for an Adaptive Flight Planner. Presented at the Aerospace Sciences Meeting, AIAA, Orlando, Fl., January 2010.

Genome sequencing will soon become a standard part of medical care, agricultural practices, drug discovery and approval, and a myriad of other applications

The Current Status and Future Outlook for Genomic Technologies

Jeffrey Fisher

Mostafa Ronaghi

Jeffrey Fisher and
Mostafa Ronaghi

Genomics emerged as a scientific field after the invention of the original DNA sequencing technique by Fredrick Sanger (Sanger et al., 1977a,b). Sanger introduced a chemical method for reading about 100 nucleotides, which, at the time, took about six months of preparation. Thanks to a large community of scientists worldwide, Sanger's technique eventually evolved to become the technology of choice for sequencing. The draft sequencing of the first human genome took about 13 years to complete, and the project cost some $3 billion.

Pyrosequencing, the second alternative technology, is based on sequencing-by-synthesis, which could be parallelized to enable higher throughput by more than 100 fold (Ronaghi et al., 1996, 1998). Pyrosequencing was used to sequence thousands of microbial and larger genomes, including James Watson's genome.

In 2006, a private company (Illumina) introduced reversible dye-terminator sequencing-by-synthesis (Bentley et al., 2006). This technology has increased throughput by ~10,000 fold in the last four years and reduced the cost of sequencing a human genome to less than $10,000. The most recent system based on this chemistry allows sequencing of several human genomes in a single run.

Jeffrey Fisher is a scientist in the Applied Genomics Group, Research and Development, and Mostafa Ronaghi is senior vice president and chief technology officer at Illumina in San Diego.

In this article, we describe dye-terminator sequencing-by-synthesis and efforts to reduce costs even further. In addition, we discuss emerging applications and challenges to bringing genomics into the mainstream.

Background

On the most fundamental level, sequencing the genome consists of just a handful of basic biochemical steps. The challenge is posed by the enormous scale of molecularly encoded information—two almost identical strands, each consisting of 3.2 billion base pairs of information for the human genome—that must be processed through those steps. Furthermore, a typical genome is read to 30X coverage, which means that each base pair is read on average 30 times (on separate strands of DNA), giving a total throughput per genome of 100 billion base pairs.

The processing and reading of these immense amounts of information has been made possible by the adoption of engineering-based approaches to massive parallelization of the sequencing reactions. All current-generation sequencing platforms coordinate chemical, engineering, and computation subsystems on an unprecedented scale (measured in information throughput) (Figure 1).

DNA Sequencing

Sequencing, which determines the arrangement of the four genetic bases (A, T, C, and G) in a given stretch of DNA, relies on four steps (Metzker, 2010; Pettersson et al., 2009; Shendure and Ji, 2008):

1. **Fragmentation**—breaking the genome into manageable segments, usually a few hundred base pairs long.
2. **Isolation**—capturing the segments in a way that keeps the signals they present distinct.
3. **Amplification**—although single-molecule techniques can theoretically proceed without this step, most systems apply some form of clonal amplification to increase the signal and accuracy of sequencing.
4. **Readout**—transforming the genetic information base by base into a machine-readable form, typically an optical (fluorescent) signal.

Although the field of genomics has evolved in recent years to include a variety of sequencing systems, including some that do not necessarily follow this exact pattern, the majority of commercial platforms use all four steps in one form or another.

The Genome Analyzer and HiSeq Systems

The Illumina Genome Analyzer and HiSeq systems are examples of the massively parallel nature of the biochemical workflow described in the four steps listed above (Bentley et al., 2008). First a sample of DNA is fragmented into segments ~400 base pairs long, and oligonucleotides of known sequence are ligated to the ends. These ligated adapters function as "handles" for each segment, allowing it to be manipulated in downstream reactions. For example, they provide a means of trapping the DNA segment in the flow cell and later releasing it. They also provide areas where primers can bind for the sequencing reaction.

Next, the sample is injected into a flow cell containing a lawn of oligonucleotides that will bind to the adapters on the DNA segments (Figure 2a). The concentration is carefully controlled so that only one strand is present in a given area of the chip—representing the signal isolation step. The segment is then amplified in place by means of a substrate-bound polymerase chain reaction process (called bridge PCR), until each single segment has grown into a cluster of thousands of identical copies of the sequence (Figure 2b). A single flow cell finally contains several hundred million individual clusters. Although they are now larger than the initial single strand, the clusters remain immobile and physically separated from each other, making it possible to visually distinguish them during the readout step.

The genetic sequence is then transformed into a visual signal by synthesizing a complementary strand,

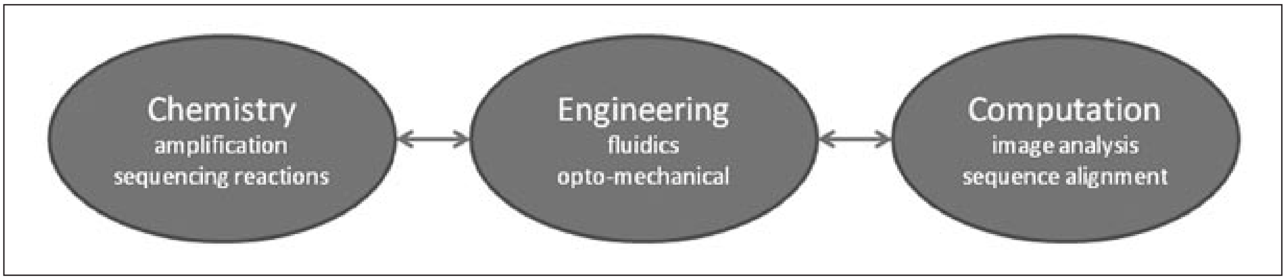


FIGURE 1 Modern sequencing requires highly coordinated subsystems to handle the throughput of massive amounts of information.

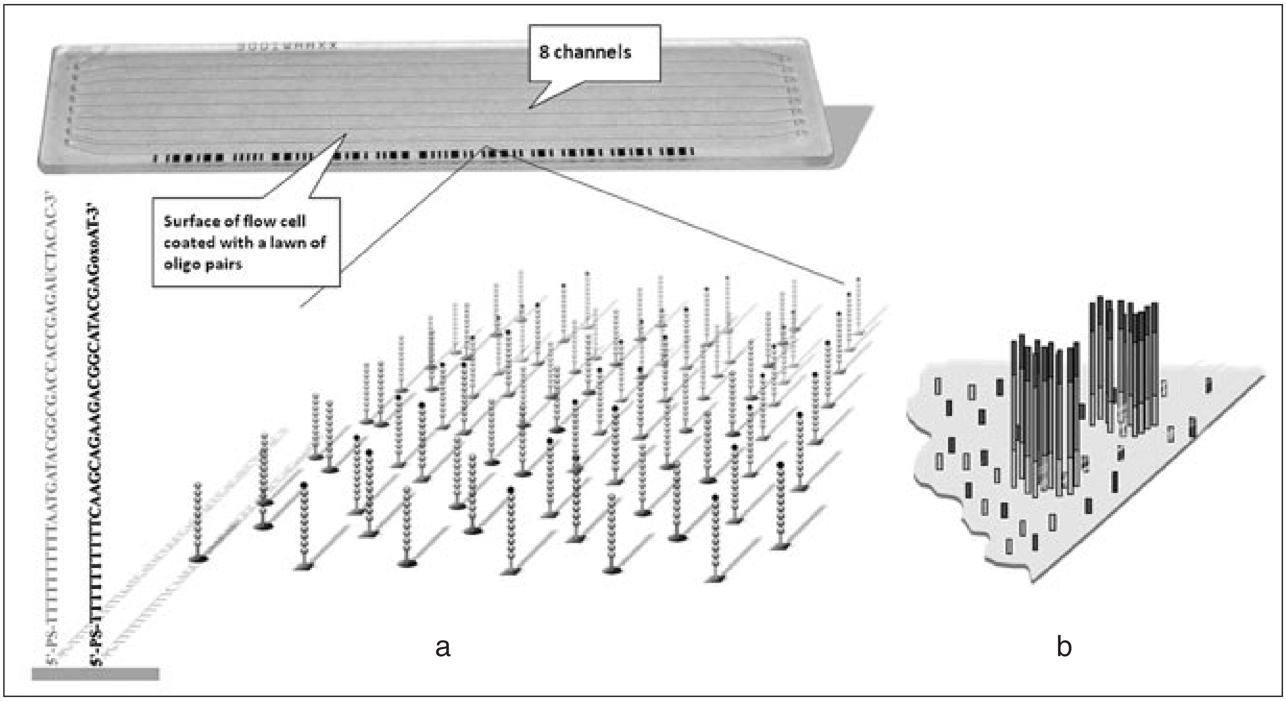


FIGURE 2 (a) Adapter-ligated segments of DNA are loaded into a flow cell coated with a lawn of oligonucleotides to capture and bind DNA segments. (b) Once attached, the DNA is PCR amplified in place to form clonal clusters.

one base at a time, using nucleotides with four separate color tags (Figure 3a). For each cycle (during which a single base per cluster is read), DNA polymerase incorporates a single nucleotide that matches the next base on the template sequence. All four nucleotides, each carrying a different dye, are added in a mixture, but only one nucleotide is incorporated into the growing DNA strand. The incorporated nucleotide has a terminator group that blocks subsequent nucleotides from being added. The entire flow cell is then imaged, and the

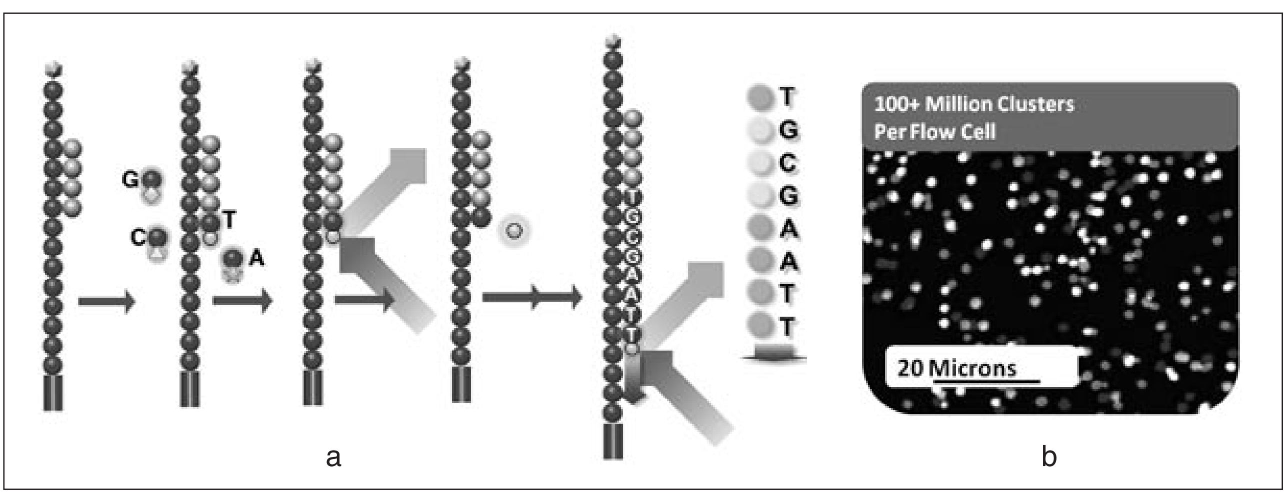


FIGURE 3 Sequencing-by-synthesis with reversible dye-terminator chemistry. During each round, (a) fluorescently-labeled nucleotides (with different colors for A, T, C, and G) are added to the flow cell. The nucleotide complementary to the next open position is then incorporated, and the entire flow cell is imaged, with the color of each cluster showing exactly which nucleotide was incorporated into that cluster during the current cycle. The label and terminator are then cleaved from the nucleotide so the reaction can begin anew. This process is repeated hundreds of times to read out a single contiguous sequence. The image map taken each cycle (b) is a four-color mosaic of clusters each emitting a color corresponding to the most recently incorporated nucleotide. The clusters are randomly distributed across the surface of the flow cell, but because they are immobilized at a fixed location, the progress of each can be followed from cycle to cycle.

color of each cluster indicates which base was added for that sequence (Figure 3b). Finally, the terminator group and fluorophore are cleaved (i.e., chemically separated) from the nucleotide, and the cycle begins again.

This process is repeated until each cluster has been read 100 to 150 times. The segment can then be "flipped over," and another 100 to 150 bases of sequence information can be read from the other end. Thus, the total amount of information that can be garnered from a single flow cell is directly proportional to the number of clusters and the read length per cluster, both of which represent targets for improvement as we continually increase system throughput.

Moore's Law and Genomics

The often-quoted Moore's law posits that the number of transistors on an integrated circuit will double every 18 to 24 months, consequently reducing the cost per transistor (Figure 4). Sequencing costs have demonstrated a similar exponential decrease over time, but at an even faster pace.

One factor that has made this possible is that, unlike transistors, which have a density limited to improvements in the two-dimensional efficiency (surface area) of the chip, sequencing density can increase along a "third dimension," which is the read length. Therefore, each subsystem in Figure 1 can be improved to increase the total throughput of the system. Improvements in the chemistry have resulted in improved accuracy, longer read lengths, and shorter cycle times. In addition, by increasing both the area of the flow cell and the density of clusters, the total number of clusters has also been increased.

The engineering subsystem has doubled the throughput by using both the top and bottom of the flow cell for cluster growth. Cluster density has been increased by improving the optics and the algorithms that detect clusters. Total run time is regularly decreased by using faster chemistries, faster fluidics, faster optical scanning, and faster algorithms for image processing and base calling. On the one hand, improvements in each subsystem independently contribute to increases in throughput. On the other hand, an improvement in one system often becomes the leading driver for advances in the others.

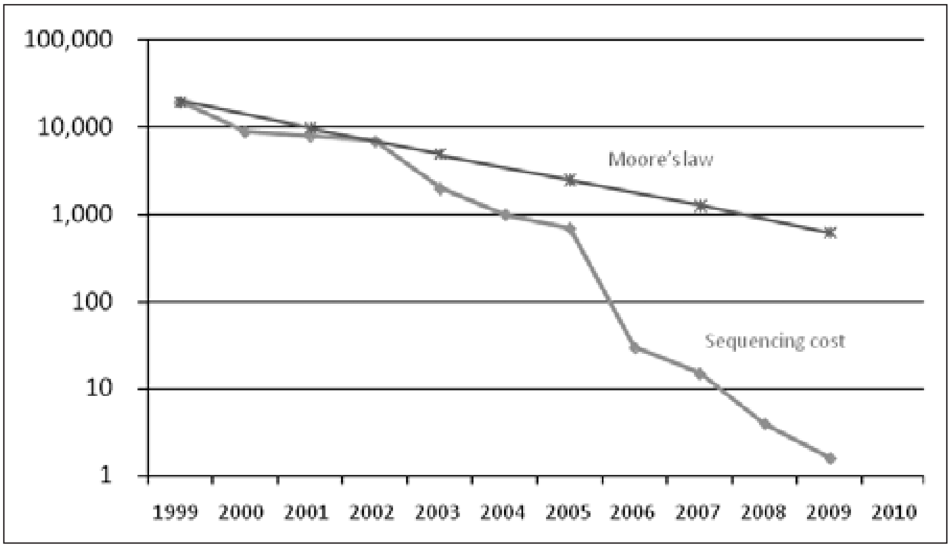


FIGURE 4 As the throughput of sequencing systems increases exponentially, the costs drop accordingly, outpacing the rate of change in Moore's law (arbitrary cost units along the y-axis).

Frontiers in Genomics

The way forward lies in improving the technology so that it can be adapted to a broader range of applications. Three ways to achieve this are: (1) increasing accuracy to enable all diagnostic applications; (2) increasing the sensitivity of the system so that it can more robustly handle lower signal-to-noise ratios; and (3) increasing throughput to drive down costs.

Improving the Accuracy of Diagnoses

Using the methods described above, one can sequence an entire human genome starting with less than one microgram of DNA, about the amount of genetic material in fewer than 150 cells. However, there are other types of samples for which even this relatively modest amount of material is difficult to come by. For example, many researchers are beginning to look at the genomics of single cells—and not just one single cell, but processing small populations individually to evaluate the heterogeneity in the group (Kurimoto and Saitou, 2010; Taniguchi et al., 2009; Walker and Parkhill, 2008). However, because a cell contains only about 6 picograms (pg) of genomic DNA and 10 pg of RNA, the corresponding signal is many orders of magnitude weaker than normal.

Another sample type that would benefit from improved assay sensitivity is a formalin-fixed, paraffin-embedded (FFPE) sample. FFPEs are histological tissue samples that have been chemically fixed to reduce degeneration, so they can be stained and examined under a microscope. Because there are huge archives of historical samples for which detailed patient outcomes are already known, researchers can use FFPE samples as resources to improve diagnosis by tracking down the genetic markers of disease. In addition, more accurate prognoses and more effective treatments are possible by studying the correlation between disease progression and genetic type in these earlier patients.

Unfortunately, fixing, staining, and storing FFPE samples can break down the genetic material, thus making sequencing or genotyping much more difficult. Nevertheless, the ability to use these samples and perform genomic analysis on them represents an invaluable resource for tracking down genetic contributions to disease and wellness (Bibikova et al., 2004; Lewis et al., 2001; Schweiger et al., 2009; Yeakley et al., 2005).

> *When whole-genome sequencing costs just hundreds of dollars, it will transform diagnosis, prognosis, drug development, agriculture, and biology.*

Increasing Sensitivity to Signal-to-Noise Ratios

In some cases, the signal itself is present at normal levels, but a much higher level of background noise drowns it out. For example, there has recently been a good deal of interest in studying the microbiome of different environments, such as soil, seawater, and the human gut (Gill et al., 2006; Turnbaugh et al., 2007; Woyke et al., 2009). In these cases, the genetic diversity of the sample can make it difficult to separate the components of different organisms.

Genomics also plays a vital role in the study of cancer (Balmain et al., 2003; Jones and Baylin, 2002; Stratton et al., 2009), which is defined by its genetic instability and pathology (Lengauer et al., 1998; Loeb, 1991, 2001). However, cells taken even from the same tumor can exhibit extreme genetic heterogeneity making increased sensitivity and detection key to distinguishing the often subtle differences that lead to one outcome as opposed to another. Sequencing this kind of sample requires much deeper coverage (7,200X read redundancy per base) than the typical 30X coverage for a homogenous sample.

Reducing Costs

Increases in throughput will affect the quantity of genetic information available, and the resultant decrease in cost will open up completely new markets, representing a qualitative shift in the ways in which genomics impacts our daily lives. When the cost of sequencing an entire genome is comparable to the current cost of analyzing a single gene, the market will experience a watershed moment as a flood of new applications for sequencing become possible. Diagnosis, prognosis, pharmacogenomics, drug development, agriculture—all will be changed in a fundamental way.

When whole-genome sequencing is priced in the hundreds of dollars, it will begin to be used all around us. It will become standard to have a copy of one's own genome. As *de novo* sequencing brings the genomes of an increasing variety of organisms into the world's databases, the study of biology will change from a fundamentally morphological classification system to genetically based classification. In agriculture, sequencing can act as an analog of a tissue-embedded radio frequency identification device (RFID); but instead of having to tag a sample with an electronic technology, we will simply extract some genetic material from a sample and sequence it, leading back to the very farm from which it came.

Today, it typically takes 12 years to bring a new drug to market, half of which is spent on discovery and half on approval; sequencing plays a role in both stages. During drug discovery, the pathways elucidated by genomic analysis lead to targeted development and shorter discovery cycles. The approval process will be facilitated by using genetic testing to define the patient populations involved in the testing of new drugs. Genetic testing will make it possible to account for genetic variation in a trial subject group when assessing efficacy and side effects. This will also lead to an improvement in treatment after a drug has been approved, as companion genetic tests for drugs will help doctors make informed decisions about how a drug might interact with a patient's genetic makeup.

Challenges

Peripheral Systems

Achieving the improvements described above will require overcoming technical obstacles directly related to chemical, engineering, and computation modules of sequencing systems. However, some of the most significant bottlenecks to throughput are found not in sequencing itself, but in the peripheral (or ancillary) systems upstream and downstream of the process.

On the upstream side, for example, the rate at which samples are sequenced now outpaces the rate at which they can be prepared and loaded. At a conference last month, the Broad Institute described its ongoing efforts to increase the number of samples a technician can prepare each week from 12 or 15 to almost one thousand by making sample preparation faster, cheaper, and with higher throughput (Lennon, 2010).

On the downstream end of the system we are beginning to bump up against throughput limits as well. Currently the HiSeq 2000 system produces about 40 GB (represented either as gigabases or gigabytes) of sequence information per day (*http://www.illumina.com/systems/hiseq_2000.ilmn*). Information generated and accumulated at that rate cannot conveniently be handled by a local desktop computer. The storage, manipulation, and analysis of this information can only be done in the "cloud," whether by local servers, dedicated off-site servers, or third parties.

Although this amount of information can easily be transferred over network hardware, as sequencing systems advance to more than 1 terabyte per day, the physical infrastructure of data networks will begin to become a limiting factor. New algorithms and standards for non-lossy compression of whole-genome data sets into files recording an individual's genetic variations (single nucleotide polymorphisms [SNPs], copy number, etc.) from the reference genome will reduce the data burden a thousand-fold. However, even solving these kinds of bandwidth issues will not address the question of how one analyzes and uses the huge amount of information being generated. To put this in perspective, a single machine can now produce the same amount of data in one week as the Human Genome Project produced in 10 years.

Non-technical Challenges

Some of the most significant challenges facing mainstream genomics are decidedly non-technical in nature. Like many information-based fields, the pace of innovation is outstripping the rate at which legislation and regulation can keep up. Laws designed prior to the genomic revolution are being shoehorned to fit technologies and situations for which there are no clear precedents.

The regulatory landscape must be more clearly defined, so companies can move forward with confidence in leveraging innovations to improve people's lives. Simultaneously, we must raise public awareness of genomic technologies to dispel myths and promote a realistic, more accurate understanding of the importance of genomics to the health of both individuals and society as a whole.

Summary

Genomics has emerged as an important tool for studying biological systems. Significant cost reductions in genomic sequencing have accelerated the adaptation of this technology for applications in a variety of market segments (e.g., research, forensics, consumer products, agriculture, and diagnostics). The most important factor in reducing cost is increasing throughput per day. We predict that the cost of sequencing an entire genome will drop to a few hundred dollars in the next few years as throughput rises with increasing density, longer read length, and shorter cycle time.

In a year or so, the cost of genome sequencing will be less than the cost of single-gene testing, which by itself has already brought significant cost savings to health care. We also predict that genome sequencing will soon become a standard part of medical practice and that in the next 15 years everybody in the Western world will be genome-sequenced.

References

Balmain, A., J. Gray, and B. Ponder. 2003. The genetics and genomics of cancer. Nature Genetics 33(Suppl): 238–244.

Bentley, D.R. 2006. Whole-genome re-sequencing. Current Opinion in Genetics and Development 16(6): 545–552.

Bentley, D.R., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456(7218): 53–59.

Bibikova, M., D. Talantov, E. Chudin, J. M. Yeakley, J. Chen, D. Doucet, E. Wickham, D. Atkins, D. Barker, M. Chee, Y. Wang, and J.-B. Fan. 2004. Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. American Journal of Pathology 165(5): 1799–1807.

Gill, S.R., M. Pop, R.T. DeBoy, P.B. Eckburg, P.J. Turnbaugh, B.S. Samuel, J.I. Gordon, D.A. Relman, C.M.

Fraser-Liggett, and K. E. Nelson. 2006. Metagenomic analysis of the human distal gut microbiome. Science 312(5778): 1355–1359.

Jones, P.A., and S.B. Baylin. 2002. The fundamental role of epigenetic events in cancer. Nature Reviews. Genetics 3(6): 415–428.

Kurimoto, K., and M. Saitou. 2010. Single-cell cDNA microarray profiling of complex biological processes of differentiation. Current Opinion in Genetics and Development 20(5): 470–477.

Lengauer, C., K.W. Kinzler, and B. Vogelstein. 1998. Genetic instabilities in human cancers. Nature 396(6712): 643–649.

Lennon, N. 2010. Optimization of Sample Preparation for Next-Generation Sequencing. Presented at the Evolution of Next-Generation Sequencing Conference, September 27–29, 2010, Providence, Rhode Island.

Lewis, F., N.J. Maughan, V. Smith, K. Hillan, and P. Quirke. 2001. Unlocking the archive–gene expression in paraffin-embedded tissue. Journal of Pathology 195(1): 66–71.

Loeb, L.A. 1991. Mutator phenotype may be required for multistage carcinogenesis. Cancer Research 51(12): 3075–3079.

Loeb, L.A. 2001. A mutator phenotype in cancer. Cancer Research 61(8): 3230–3239.

Metzker, M.L. 2010. Sequencing technologies—the next generation. Nature Reviews. Genetics 11(1): 31–46.

Pettersson, E., J. Lundeberg, and A. Ahmadian. 2009. Generations of sequencing technologies. Genomics 93(2): 105–111.

Ronaghi, M., S. Karamohamed, B. Pettersson, M. Uhlén, and P. Nyrén. 1996. Real-time DNA sequencing using detection of pyrophosphate release. Analytical Biochemistry 242(1): 84–89.

Ronaghi, M., M. Uhlén, and P. Nyrén. 1998. A sequencing method based on real-time pyrophosphate. Science 281(5375): 363, 365.

Sanger, F., G.M. Air, B.G. Barrell, N.L. Brown, A.R. Coulson, C.A. Fiddes, C.A. Hutchison, P.M. Slocombe, and M. Smith. 1977a. Nucleotide sequence of bacteriophage phi x174 DNA. Nature 265(5596): 687–695.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977b. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74(12): 5463–5467.

Schweiger, M.R., M. Kerick, B. Timmermann, M.W. Albrecht, T. Borodina, D. Parkhomchuk, K. Zatloukal, and H. Lehrach. 2009. Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. PLoS ONE 4(5): e5548.

Shendure, J., and H. Ji. 2008. Next-generation DNA sequencing. Nature Biotechnology 26(10): 1135–1145.

Stratton, M.R., P.J. Campbell, and P.A. Futreal. 2009. The cancer genome. Nature 458(7239): 719–724.

Taniguchi, K., T. Kajiyama, and H. Kambara. 2009. Quantitative analysis of gene expression in a single cell by QPCR. Nature Methods 6(7): 503–506.

Turnbaugh, P.J., R.E. Ley, M. Hamady, C.M. Fraser-Liggett, R. Knight, and J.I. Gordon. 2007. The human microbiome project. Nature 449(7164): 804–810.

Walker, A., and J. Parkhill. 2008. Single-cell genomics. Nature Reviews Microbiology 6(3): 176–177.

Woyke, T., G. Xie, A. Copeland, J.M. González, C. Han, H. Kiss, J.H. Saw, P. Senin, C. Yang, S. Chatterji, J.-F. Cheng, J.A. Eisen, M.E. Sieracki, and R. Stepanauskas. 2009. Assembling the marine metagenome, one cell at a time. PLoS One 4(4): e5299.

Yeakley, J.M., M. Bibikova, E. Chudin, E. Wickham, J.B. Fan, T. Downs, J. Modder, M. Kostelec, A. Arsanjani, and J. Wang-Rodriguez. 2005. Gene expression profiling in formalin-fixed, paraffin-embedded (ffpe) benign and cancerous prostate tissues using universal bead arrays. Journal of Clinical Oncology 23(16S): 9526.

Microscopic systems with complex functionalities have been the inspiration for the emerging field of synthetic biology.

Autonomous Systems and Synthetic Biology

Henry Hess is associate professor, Department of Biomedical Engineering, Columbia University.

Henry Hess

Autonomous systems have helped solve a variety of engineering challenges, from drastic changes in manufacturing processes since the 1950s to the exploration of space and oceans. Recently, autonomous systems conceived at the micro- and nanoscale levels are being used to address challenges in materials and biomedical applications. These microscopic devices and their applications are inspired by autonomous biological systems that operate both individually and collectively.

A prototypical autonomous biological device is a *Vibrio cholerae* bacterium that packs the ability to move, sense, target, adapt, and release active substances into just a few cubic centimeters. An example of a smart biological material, muscle, which is composed of autonomous systems, is hierarchically assembled from microscopic subunits, has the ability to exchange information with the environment via electrical and mechanical stimuli, and incorporates energy-conversion modules and self-healing abilities.

Both *V. cholerae* bacteria and muscle cells are significant achievements in engineering by evolution. Both can operate in diverse environments with low power consumption and limited computing resources.

These and other examples of microscopic systems with complex functionalities have been the inspiration for the emerging field of synthetic biology. A prominent research strategy in this new field—inspired by the successful large-scale integration of electronic circuits—is to focus on

the design of standardized gene circuits that can serve as modules of complex programs to be executed by bacterial cells. This approach is akin to the delivery of a set of well organized blueprints to a contract manufacturer who then manufactures equipment according to the delivered specifications, learns from the experience of technicians to operate the equipment, and produces a product of interest.

A second strategy, which is the topic of this paper, is to develop the technical expertise to rationally design complex, interacting microscopic systems (Schwille and Diez, 2009). This approach builds on nanotechnology, as well as on increasingly complex *in vitro* experiments in cell biology. In these experiments, we replicate critical cellular functions to test our understanding of essential and auxiliary biological mechanisms and components.

The challenges in designing biomimetic systems using nanoscale building blocks include (1) controlling their operation in the presence of Brownian motion and other sources of noise, (2) integrating molecular information properly, (3) addressing lifetime and reliability issues, and (4) anticipating and using emergent phenomena.

Kinesin-Powered Molecular Shuttles

Kinesin motors are proteins that use ATP molecules as fuel to generate mechanical work (Howard, 2001). A microtubule assembled from thousands of tubulin proteins serves as a track for the kinesin motor. For each ATP molecule a kinesin motor hydrolyzes, it takes one step (i.e., moves 8 nanometers [nm]) along the microtubule track. The kinesin motor can advance against a force of about 5 piconewtons (pN), in the process converting more than 50 percent of the free energy of ATP hydrolysis into mechanical work.

Within cells, kinesin is primarily responsible for transporting molecular cargo from the center of the cell to the periphery. Biophysicists have developed the ability to observe and manipulate kinesin motors and the associated microtubule filaments outside the cell in so-called in vitro gliding assays. In these assays, kinesins are adhered to a surface, and fluorescently labeled microtubules are propelled by kinesin motors in the presence of ATP (Figure 1).

In addition to enabling the study of motor proteins, microtubules propelled by kinesin motors serve as a nanoscale transport system. By controlling the direction of the microtubules, the attachment and detachment of cargo to microtubules, and the supply of ATP fuel, microtubules can be induced to act as nanoscale delivery trucks, or "molecular shuttles" (Hess and Vogel, 2001). Assembled from biological components with unmatched functionality, these molecular shuttles can be used to explore design concepts for nanoscale systems and devices (Hess et al., 2002a,b).

Although individual molecular shuttles can be controlled (van den Heuvel et al., 2006), the inherent advantages of molecular devices can be better exploited by putting aside costly efforts to achieve individual control and accepting instead the autonomous operation of molecular shuttles in an externally directed "swarm."

A suitable analogy to this scenario is an anthill. Although in principle it is possible to induce an individual ant to perform a specific task, the anthill, as a

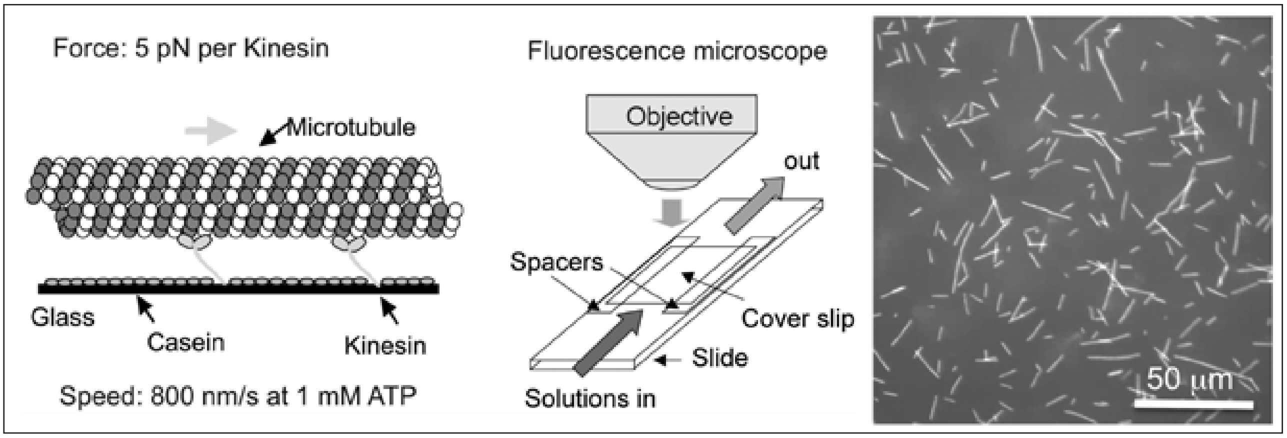


FIGURE 1 Surface-adhered kinesin motors can propel fluorescently labeled microtubules (diameter 25 nm) across a surface. A coating of casein proteins on the surface prevents the kinesin motor domains from attaching to the surface. The experiment is conducted in a cell composed of a cover slip, spacers, and a slide. When microtubule motion is observed with a fluorescence microscope, microtubules appear as fluorescent rods, and the kinesin motors on the surface are invisible.

complex system, relies on the emergence of useful actions that result from the autonomous decisions of individual ants. Figure 2 shows the assembly of "nanospools" from "sticky" microtubules, a striking example of emergence in swarms of molecular shuttles (Hess et al., 2005).

Mechanical Engineering at the Interface of Biophysics and Supramolecular Chemistry

Precise measurements of nanometer lengths and molecular arrangements and interactions are the building blocks for the rational engineering of molecular shuttles. For example, fluorescence-interference contrast microscopy enables the measurement of the 20 nm "ground clearance" of molecular shuttles (Kerssemakers et al., 2009). Other experiments have determined the distribution of cargo-binding linker molecules on the shuttles and elucidated the complex, glue-like interaction between complementary linker molecules on a shuttle and its cargo (Pincet and Husson, 2005).

The combination of precise spatial information and detailed knowledge of molecular interactions enables us to predict emerging properties, such as an optimal velocity for cargo loading (Agarwal et al., 2009). This optimal velocity is a result of (1) the milliseconds required for the glue-like interaction between biotin and streptavidin to strengthen and (2) the velocity-dependence of the number of attempts to form such bonds.

The importance of such studies is that they transition from the reductionist analysis of molecular processes often practiced in biophysics and supramolecular chemistry to an engineering analysis of emerging phenomena resulting from the assembly of well understood building blocks into complex and artificial structures.

The challenge is made more daunting because classic engineering tools, such as technical drawings, have not yet been adapted to capture the dynamic nature of molecular structures. Questions such as whether the fluctuating reach of a linker molecule should be represented by its most likely or it most extended configuration (or both) are very important in a complex technical drawing (Figure 3) showing molecules of different sizes and flexibilities interacting on a wide range of time scales.

"Smart Dust" Biosensors as Applications of Molecular Shuttles

Microorganisms, which have the ability to detect a variety of analytes with high specificity and sensitivity, process the incoming information, and communicate their measurements, excel at biosensing. From the perspective of an engineer, microorganisms act as microscopic sensor packages that are immersed into the sample (their environment) and collectively respond to analytes.

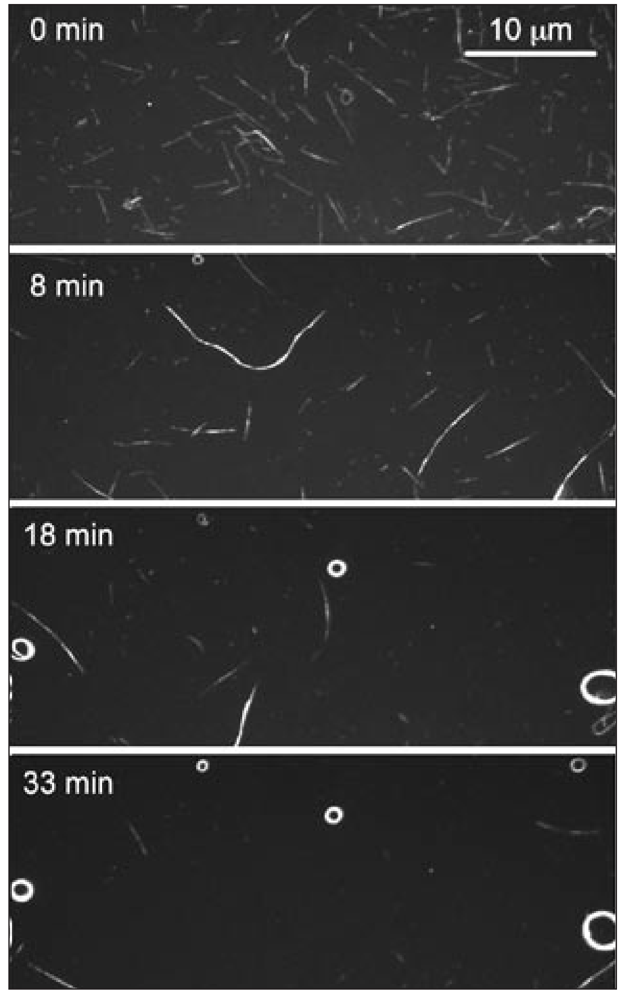


FIGURE 2 Biotin-functionalized microtubules rendered "sticky" by a partial coating of streptavidin self-assemble into nanowires and ultimately into nanospools.

This concept is transferred to the engineering domain by "smart dust," which is an attempt to create highly integrated microscopic sensors in large numbers for remote detection scenarios (Kahn et al., 2000; Sailor and Link, 2005). With support from the DARPA Biomolecular Motors Program, five research teams working collaboratively pursued the creation of "smart dust" biosensors, whose core components are molecular shuttles (Bachand et al., 2009). In these sensors, antibody-functionalized molecular shuttles capture analytes, tag them with fluorescent particles, and transport them to a deposition zone for detection (Figure 4).

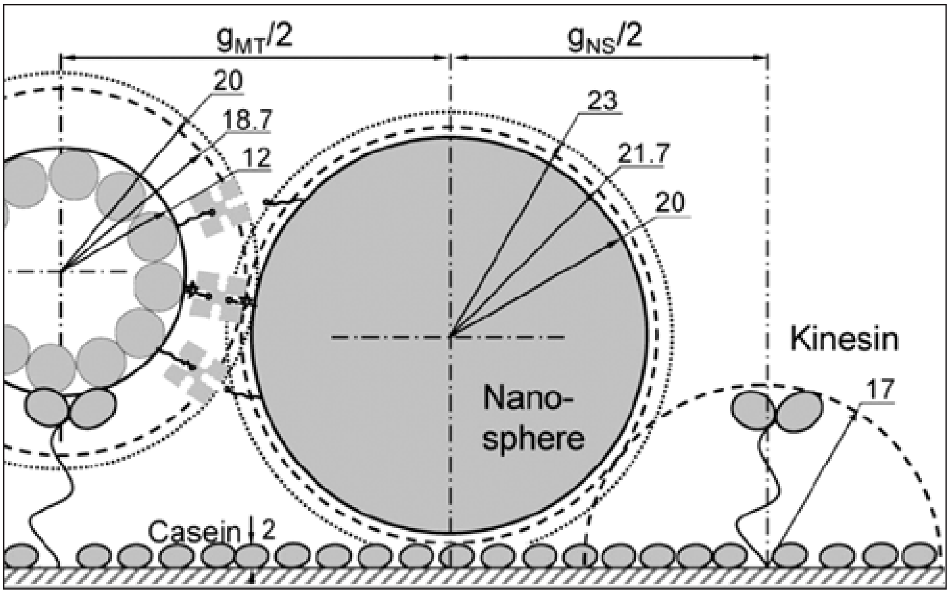


FIGURE 3 Understanding the interaction between a gliding microtubule functionalized with linker molecules (left) and a surface-adhered nanosphere coated with the complementary molecules requires a detailed analysis of the spatial arrangement. Lengths are measured in nanometers, dashed lines represent most likely positions, and dotted lines represent maximal extensions of biotin linkers and kinesin molecules. Adapted from Agarwal et al., 2009.

This design has several "biomimetic" aspects. First, the concept of smart dust per se is inspired by biological organisms. Second, the components of the molecular shuttles—kinesin motors, microtubules, and antibodies—are biological in origin. And third, the principle of operation, a swarm of unsophisticated, autonomous devices creating a detectable signal, is also bio-inspired.

The disadvantages of such a hybrid device are also closely related to its biological origins. Although maintaining the biological components in a relatively narrow temperature range at a given pH and with well defined buffer conditions is feasible in the laboratory, it cannot be guaranteed in the field. Therefore, although strategies for extending the lifetime and optimizing storage conditions have been explored (Boal et al., 2006; Seetharam et al., 2006), the fragility of the biological components is a key shortcoming. Successful proof-of-principle demonstrations of application concepts with hybrid devices will therefore require the development of synthetic components, such as synthetic molecular motors (Kay et al., 2007).

However, even when synthetic devices come close to achieving the performance of their natural counterparts, there is still an open question about whether there will be a fundamental trade-off between robustness and performance that will lead to similarly fragile synthetic devices. In addition, we do not know if biomolecular motors represent the pinnacle of achievable performance in terms of force, power, and energy

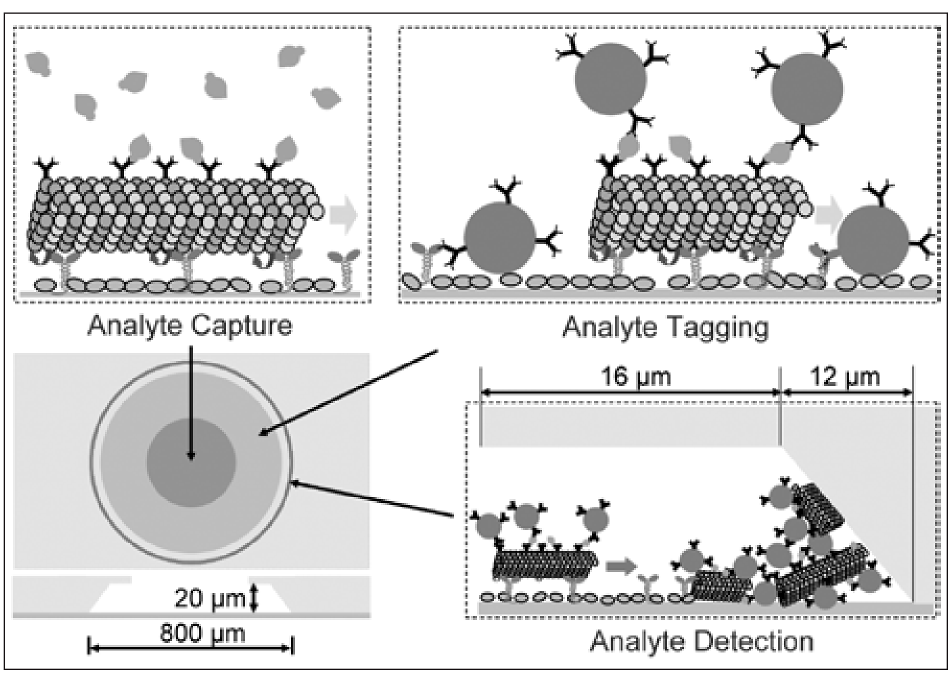


FIGURE 4 A "smart dust" biosensor based on molecular shuttles. Antibody-functionalized microtubules capture analyte molecules, such as protein biomarkers, in the center of a circular well and transport them across the surface. Collisions with antibody-coated fluorescent particles leads to particle capture if analyte is present. Eventually, shuttles reach the periphery of the well, where they accumulate and their cargo of analytes and fluorescent particles can be detected optically.

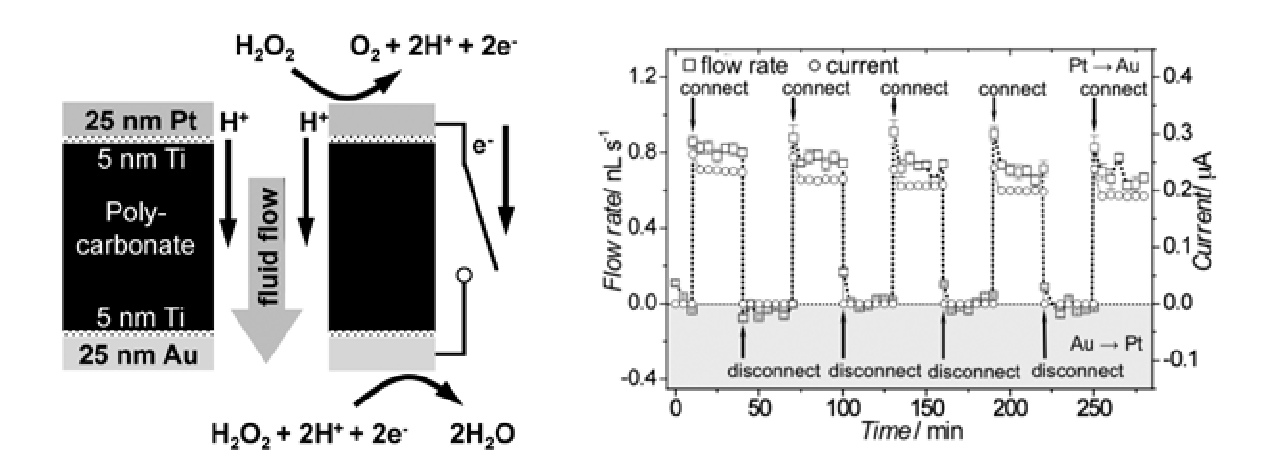


FIGURE 5 A biomimetic, self-pumping membrane (Jun and Hess, 2010). A thin plastic membrane with 1 μm diameter straight channels is coated with a thin film of platinum on one side and a thin film of gold on the other. When the membrane is submerged in a hydrogen peroxide solution and the two electrodes are electrically connected, the system acts as an integrated fuel cell and electroosmotic pump that uses the chemical energy of the hydrogen peroxide to power fluid flow across the membrane.

efficiency, or if the use of synthetic materials and fabrication techniques will result in dramatic improvements in one or more metrics.

Transitioning to Synthetic Materials

In addition to the exciting research and potential applications of hybrid systems mentioned above, the performance and capabilities of synthetic systems are being steadily improved by ongoing efforts in the semiconductor industry to miniaturize energy sources, sensors, computing elements, and communications systems. This technological trend may enable the design of "nanomorphic cells" (Cavin and Zhirnov, 2008). The creation of entirely new classes of autonomous microdevices with applications, for example, as smart therapeutic systems in medicine would significantly benefit the semiconductor industry.

Although a number of interesting avenues will be explored in the design of nanomorphic cells, my collaborators and I are primarily interested in providing mechanical work for locomotion using chemical fuel sources. The conversion of chemical energy into mechanical work with synthetic nanoscale devices is still in its infancy. The design of complex organic molecules capable of contraction or rotation when provided with fuel molecules has made significant progress but has not yet resulted in designs that can compete with biomolecular motors (Kay et al., 2007).

One successful approach has been to mimic bacterial motility by platinum-gold nanorods in a hydrogen peroxide solution (Hong et al., 2007). The nanorods catalyze the decomposition of the hydrogen peroxide, which in turn propels them forward. Surprisingly, the nanorods have a distinctly "chemotactic" response, moving toward higher hydrogen peroxide concentrations. The analysis of the process has informed our understanding of the mechanism supporting chemotaxis in bacteria.

On the basis of a similar combination of a "compartmentless" fuel cell and electroosmotic pumping, we were able to engineer a fully synthetic membrane that mimics the ability of cellular membranes to actively transport solutes using chemical energy harvested from the solution (Jun and Hess, 2010). A platinum electrode and a gold electrode on the surface of a polycarbonate membrane were electrically connected. When placed in a hydrogen peroxide solution, the fluid was pumped through the membrane (Figure 5).

In a sense, our efforts to create biomimetic functional materials and bio-inspired nanodevices are following the arc of development of human flight. First came the study of biological systems. This was followed by the creation of hybrid systems to elucidate the key principles of flight. Ultimately, synthetic flying machines were developed.

Conclusions

The engineering of molecular shuttles and other autonomous systems is an exciting aspect of nanotechnology in which progress relies heavily on the use of biological components. The assembly of biological building blocks into newly designed structures is a very stringent test of our assumptions about biological

nanomachines and their interactions. As Richard Feynman said, “What I cannot create, I do not understand.” The design process forces us to see biology through the eyes of an engineer and, in the process, to ask about friction, wear, fatigue, and other quintessential engineering questions.

References

Agarwal, A., P. Katira, and H. Hess. 2009. Millisecond curing time of a molecular adhesive causes velocity-dependent cargo-loading of molecular shuttles. Nano Letters 9(3): 1170–1175.

Bachand, G.D., H. Hess, B. Ratna, P. Satir, and V. Vogel. 2009. “Smart dust’” biosensors powered by biomolecular motors. Lab on a Chip 9(12): 1661–1666.

Boal, A.K., H. Tellez, S.B. Rivera, N.E. Miller, G.D. Bachand, and B.C. Bunker. 2006. The stability and functionality of chemically crosslinked microtubules. Small 2(6): 793–803.

Cavin, R., and V. Zhirnov. 2008. Morphic architectures: atomic-level limits. Materials Research Society Symposium Proceedings 1067(Spring, Symposium B): B1001–B1002.

Hess, H., and V. Vogel. 2001. Molecular shuttles based on motor proteins: active transport in synthetic environments. Reviews in Molecular Biotechnology 82(1): 67–85.

Hess, H., J. Clemmens, J. Howard, and V. Vogel. 2002a. Surface imaging by self-propelled nanoscale probes. Nano Letters 2(2): 113–116.

Hess, H., J. Howard, and V. Vogel. 2002b. A Piconewton forcemeter assembled from microtubules and kinesins. Nano Letters 2(10): 1113–1115.

Hess, H., J. Clemmens, C. Brunner, R. Doot, S. Luna, K.H. Ernst, and V. Vogel. 2005. Molecular self-assembly of “Nanowires” and “Nanospools” using active transport. Nano Letters 5(4): 629–633.

Hong, Y., N.M.K. Blackman, et al. 2007. Chemotaxis of nonbiological colloidal rods. Physical Review Letters 99(17): 178103.

Howard, J. 2001. Mechanics of Motor Proteins and the Cytoskeleton. Sunderland, Mass.: Sinauer.

Jun, I.K., and H. Hess. 2010. A biomimetic, self-pumping membrane. Advanced Materials (in press).

Kahn, J.M., R.H. Katz, Y.H. Katz, and K.S.J. Pister. 2000. Emerging challenges: mobile networking for “smart dust.” Journal of Communications and Networks 2(3): 188–196.

Kay, E.R., D.A. Leigh, and F. Zerbetto. 2007. Synthetic molecular motors and mechanical machines. Angewandte Chemie (International Edition) 46(1–2): 72–191.

Kerssemakers, J., L. Ionov, U. Queitsch, S. Luna, H. Hess, and S. Diez. 2009. 3D nanometer tracking of motile microtubules on reflective surfaces. Small 5(15): 1732–1737.

Pincet, F., and J. Husson. 2005. The solution to the streptavidin-biotin paradox: the influence of history on the strength of single molecular bonds. Biophysical Journal 89(6): 4374–4381.

Sailor, M.J., and J.R. Link. 2005. “Smart dust”: nanostructured devices in a grain of sand. Chemical Communications (11): 1375–1383.

Schwille, P., and S. Diez. 2009. Synthetic biology of minimal systems. Critical Reviews in Biochemistry and Molecular Biology 44(4): 223–242.

Seetharam, R., Y. Wada, S. Ramachandran, H. Hess, and P. Satir. 2006. Long-term storage of bionanodevices by freezing and lyophilization. Lab on a Chip 6(9): 1239–1242.

van den Heuvel, M.G.L., M.P. De Graaff, and C. Dekker. 2006. Molecular sorting by electrical steering of microtubules in kinesin-coated channels. Science 312(5775): 910–914.

NAE News and Notes

NAE Newsmakers

Five NAE members were named to the **Secretary of Energy's Advisory Board** (SEAB). They are **Norman R. Augustine**, retired chairman and CEO, Lockheed Martin, and former Under Secretary for the Army; **Nicholas M. Donofrio**, IBM Fellow Emeritus and former executive vice president of Innovation and Technology, IBM; **Charles O. Holliday Jr.**, chairman, Bank of America, and former CEO of Dupont; **William J. Perry**, Michael and Barbara Berberian Professor, Stanford University, and former Secretary of Defense; and **Arthur H. Rosenfeld**, Senior Faculty Scientist, University of California Lawrence Berkeley National Laboratory, and former commissioner, California Energy Commission. SEAB, which was decommissioned by the Bush administration, has been reestablished under the Federal Advisory Committee Act. The 12-member board provides advice and recommendations on the Department of Energy's basic and applied research, economic and national security policy, educational issues, operational issues, and other activities as directed by the Secretary of Energy.

On September 6, 2010, the **Kavli Prize Science Forum** was held in Oslo, Norway. This biennial international forum facilitates high-level, global discussions of major topics in science and science policy. The theme this year was "The Role of International Cooperation in Science." NAE member **John P. Holdren**, Assistant to the President for Science and Technology and director of the White House Office of Science and Technology Policy, gave the keynote address, "Climate Change Science and Policy: What Do We Know? What Should We Do?" NAE President **Charles M. Vest** moderated the panel discussion that followed.

Rafael L. Bras, provost and executive vice president for academic affairs, and K. Harrison Brown Family Chair, Georgia Institute of Technology, has been named the recipient of the **2010 Anthony J. Drexel Exceptional Achievement Award**, given by Drexel University. The $50,000 award is given in recognition of collaborative, multidisciplinary research that has made a difference in people's lives. Dr. Bras' research in the civil engineering subfields of hydrology and hydroclimatology have improved the description and forecasting of floods and precipitation. He has also made major contributions to determining the impact of deforestation on the hydrologic cycle and the evolution of landscapes under different climatic forcings and climatic disturbances.

The **National Intelligence Distinguished Service Medal** is given in recognition of "sustained selfless service of the highest order and/or extraordinary and long lasting contributions to the Intelligence Community and the United States by an individual in a position of great responsibility." In September 2010, the Director of National Intelligence, The Honorable James R. Clapper, presented the award to **George R. Cotter** "for contributions in the area of high-end computing and advanced information technology architectures."

L. Eric Cross, Evan Pugh Professor Emeritus of Electrical Engineering, Pennsylvania State University, is the recipient of the **2010 Von Hippel Award** from the Materials Research Society. The Von Hippel Award, the society's highest honor, is "conferred annually to an individual in recognition of the recipient's outstanding contribution to interdisciplinary research on materials." A founding member of the Penn State Materials Research Laboratory, Dr. Cross continues to contribute to the field of ferroelectric materials through his work on flexoelectric composites, which could make a new generation of lead-free transducers possible for use in multiple industries worldwide. He is being recognized "for his imposing leadership in the science and applications of ferroelectric materials."

Delbert E. Day, Curators' Professor Emeritus of Materials Science and Engineering and chief investigator, Missouri University of Science and Technology, has been named the **2010 Phoenix Award Glass Person of the Year**, the glass industry's top honor. The Phoenix Award is given annually to a living person who has made outstanding contributions to the industry. Dr. Day received the award on September 17 at a banquet in his honor. Dr. Day's work was critical to the development of radioactive glass microspheres that are used at more than 100 sites around

the world to treat patients with inoperable liver cancer.

Laura M. Haas, IBM Fellow and director of computer science, IBM Almaden Research Center, received the **Anita Borg Social Impact Award** from the Anita Borg Institute for Women and Technology. The purpose of the award is to honor an individual or team that has facilitated a crucial link between technology and the lives of women and society or has enabled women to have a significant impact on the design and use of technology. Dr. Haas directs research in information management, human-computer interaction, theoretical foundations of computing, and health care informatics. She is also one of the founders of IBM's information integration solutions, which today include products that can integrate both structured and unstructured data via federation, consolidation, and search.

John L. Hennessy, president, Stanford University, is the recipient of the **Dr. Morris Chang Exemplary Leadership Award**, presented by the Global Semiconductor Alliance (GSA). Presented to Dr. Hennessey on December 9 during the GSA Awards Dinner Celebration, the award is given in recognition of individuals whose exceptional contributions, vision, and global leadership have transformed and elevated the entire semiconductor industry.

Arthur H. Heuer, Distinguished University Professor and Kyocera Professor of Ceramics, Department of Materials Science and Engineering, Case Western Reserve University, was **elected a Fellow of the Materials Research Society** (MRS). MRS recognizes outstanding members whose sustained and distinguished contributions to the advancement of materials research have been internationally recognized. Dr. Heuer was honored "in recognition of outstanding and sustained contributions to several materials science subfields, including transmission electron microscopy of ceramics, transformation toughening in ceramics, dislocations in oxides, material science of microelectromechanical systems, and interstitial hardening of stainless steels."

Teresa H. Meng, Reid Weaver Dennis Professor, Department of Electrical Engineering and Computer Science, Stanford University, has been **elected an Academician of the Academia Sinica**, the highest honor given to scientists in Taiwan. Dr. Meng's work in wireless communications and digital signal processing has contributed to the proliferation of Wi-Fi devices and networks around the world. Founded in 1928, Academia Sinica is the most prestigious academic institute in Taiwan.

James A. Sethian, professor, Department of Mathematics, University of California, Berkeley, will receive the **2011 Pioneer Prize** "for pioneering work introducing applied mathematical methods and scientific computing techniques to an industrial problem area or a new scientific field of applications." Dr. Sethian pioneered mathematical and computational frameworks to track moving interfaces and has introduced these techniques into diverse areas of science and engineering. His methods have been applied to understanding the manufacture of semiconductors, detecting anatomical structures in medical images, accurately modeling combustion processes, improving industrial inkjet printing, and reconstructing geological structures from reflected waves.

Man-Chung Tang, chairman of the board, T.Y. Lin International, has been awarded the International Bridge Association for Bridge and Structural Engineering's (IABSE) **International Award of Merit in Structural Engineering** "for blending art and engineering, and successfully creating innovative concepts for signature bridges that are admired by both his peers and the general public alike." Established in 1976, the award is conferred for outstanding contributions in the field of structural engineering, with special reference to their usefulness to society. Jacques Combault, president of IABSE, presented the award to Dr. Tang on September 22, 2010, at the opening ceremony of the 34th IABSE Symposium in Venice, Italy.

The Potomac Institute for Policy Studies presented **Jeffrey Wadsworth**, president and chief executive officer of Battelle, with the Institute's **Navigator Award** at a banquet at the Willard Hotel, Washington, D.C., on October 21, 2010. The Navigator Award is presented to members of Congress and representatives of the Executive Branch, industry, and academia in recognition of distinguished contributions in the arena of science and technology.

Highlights of the 2010 NAE Annual Meeting

NAE Class of 2010.

NAE members, foreign associates, and guests gathered in Washington, D.C., this October for the 2010 NAE Annual Meeting. The meeting began on Saturday afternoon, October 2, with an orientation session for new members. That evening, the NAE Council dinner was held in the Crystal Room of the Willard Hotel in honor of the 68 new members and 9 new foreign associates.

NAE Chair **Irwin M. Jacobs** opened the public session on Sunday, October 3, with brief remarks about the impact of rapidly developing, often cross-disciplinary technologies in both traditional and newer engineering fields. He also stressed the need for transforming K–12 education, making technology available, and encouraging the study of science, technology, engineering, and mathematics (STEM) (see p. 61).

NAE president **Charles M. Vest** then delivered his annual address, "Technology and the Future of U.S. Competiveness: Nightmares and Dreams" (see pg. 63). The induction of the NAE Class of 2010 followed, with introductions by NAE Executive Officer **Lance Davis**.

The program continued with the presentation of the 2010 Founders Award and Arthur M. Bueche Award. The recipient of the 2010 Founders Award was Dr. **Robert Langer**, David H. Koch Institute Professor, Massachusetts Institute of Technology, who was honored "for the invention, development, and commercialization of methods and materials for drug delivery and tissue engineering, mentoring of scientists, and the promotion of the nation's health" (see pg. 69). **Anita K. Jones**, University Professor Emerita, University of Virginia, was presented with the 2010 Arthur M. Bueche Award. Dr. Jones was cited for "leadership in the development of U.S. science and technology policy and the development of technologies for national security, including technical contributions to high performance computing and cybersecurity" (see pg. 71).

After the awards ceremony, the Honorable **Donald C. Winter**, chair of the Deepwater Horizon Investigation Committee, shared some observations about the work of his committee.

Following a break, Dr. Vest introduced the Armstrong Endowment for Young Engineers–Gilbreth Lecturers, outstanding young engineers who have participated in NAE Frontiers of Engineering symposia. J-Y. Christine Chiu, lecturer, Department of Meteorology, University of Reading, spoke on "Climate Observation from the Atmospheric Radiation Measurement (ARM) Program." David L. Sedlak, professor, Department of Civil and Environmental Engineering, University of California at Berkeley, spoke on "Reinventing Urban Water Systems."

Two guest speakers followed. The first talk, by **N.R. Narayana Murthy**, founder-chairman and chief mentor, Infosys Technologies Inc., was entitled "Engineering in a Flat World." The second guest speaker was **John P. Holdren**, Assistant to the President for Science and Technology, and director, White

Celebrating milestones as NAE members. Left to right: Erich Block (1980), Sheila Widnall (1985), Ronald Latanision (1985), Alfred Y. Cho (1985), Richard J. Goldstein (1985), Ponisseril Somasundaran (1985), and Karl Pister (1980).

House Office of Science and Technology Policy. The title of his talk was "Meeting the Climate-Change Challenge: What Do We Know? What Should We Do?" The day ended with a reception for members and guests.

On Monday morning, at the Annual Business Session for members, Dr. Vest provided a quick summary of the organizational structure and finances of NAE and its relationship with the other members of the National Academies—the National Academy of Sciences, Institute of Medicine (IOM), and National Research Council (NRC). He highlighted NAE's need for independent funds and thanked members for their personal philanthropy and gift giving to NAE. In an overview of NAE program activities, he highlighted joint activities with IOM, the ongoing Grand Challenges Project, NAE Frontiers of Engineering activities, and the recent Frontiers of Engineering Education Symposium. He also noted NAE's significant contributions to the NRC study of the Deepwater Horizon catastrophe.

The business session was followed by the annual NAE Forum, "Global Technology: Change and Implications." Dr. Vest moderated the discussion, which focused on the many aspects of global technology and the opportunities and responsibilities of engineering leaders. The panelists were Esko Aho, former prime minister of Finland (currently executive vice president of corporate relations and responsibility, Nokia); **Bernard Amadei**, founder of Engineers Without Borders; John Seely Brown, visiting scholar, University of Southern California, self-proclaimed "Chief of Confusion," and former chief scientist of Xerox PARC; **Ruth David**, president and CEO of ANSER (Analytic Services Inc.); Eric Haseltine, consultant, former associate director for science and technology in the Office of the Director of National Intelligence, and former head of research/development at Disney Imagineering; Nicholas Negroponte, founder, One Laptop per Child, and founder and Chairman Emeritus of

Forum Panel. Left to right: Esko Aho, Raymond S. Stata, John Seely Brown, Charles Vest, Nicholas Negroponte, Ruth A. David, Eric C. Haseltine, and Bernard Amadei.

the MIT Media Lab; and **Raymond S. Stata**, cofounder and chairman of the board, Analog Devices. The forum was recorded and is available on the NAE web site (*www.nae.edu*).

On Monday afternoon, members and foreign associates participated in NAE section meetings at the Keck Center and the JW Marriott. The final event was the annual reception and dinner dance later that evening, also at the JW Marriott. Dance music was provided by the Odyssey Band.

The next annual meeting is scheduled for October 16 and 17, 2011.

Remarks by NAE Chair Irwin Jacobs

Irwin Jacobs

It is a pleasure to welcome all of you to this special evening and to congratulate the class of 2010 on behalf of the NAE Council. I believe you will find being a member of the National Academy of Engineering very rewarding, and I hope you participate in many NAE activities, helping our nation to advance by providing good advice to our government and agencies.

One of the things I enjoy each year is reading through the citations of the new members, recognizing the breadth and impact of engineering today. This is, indeed, a special time to be an engineer. So much is happening, both in traditional areas of engineering and in newer and rapidly developing fields, many of them cross-disciplinary.

In my own field, digital communications, and in wireless communications and cell phones, tremendous progress has been made. It has been fascinating to see how quickly theoretical work, both from universities and from industry research labs, is being translated into actual equipment, new forms of radio links, and new types of devices. In fact, one rationale for my decision to leave academic life after 13 years and enter industry was to help demonstrate that the theory we were teaching had significant practical applications. As many of you know, that significance is often questioned.

At present, there are about five billion subscribers on cellular systems, and just about everybody on the globe will soon have wireless voice and Internet access. That leaves much room for new device capabilities and new applications, not just in developed countries, but also many opportunities in developing countries.

It has been very exciting to introduce powerful computers that occupy only part of one semiconductor chip in a phone, run on a battery, and yet deliver computing power similar to the power of a desktop just a short while ago, and the computing power and energy efficiency continue to improve. These chips also have global positioning system, camera, video, television, 2- and 3-D display support, and many other capabilities. Now it is possible for young engineers to develop all types of new applications that can really impact people's lives.

New devices impact people in all walks of life, including fishermen and farmers. One interesting service in India provides information to fishermen about the weather, where higher priced fish might be biting, and where they can land fish and get the best price. This information eliminates some of the middlemen and allows people to pursue better lives.

This type of assistance might have another important impact. By noticing the advantages, but also the limitations, of obtaining information via voice and simple pictures, people may be inspired to become more literate, to gain access to additional useful information through reading. So, technology has many different impacts around the world.

Engineering has also led to significant changes in traditional products—automobiles, for example, are going through major changes in response to concerns about the environment and energy. We are seeing the rapid development and introduction of hybrid cars, electric cars, and perhaps cars with fuel cells that run on hydrogen.

Interestingly, there are also continuing changes in cars with internal combustion engines. I just attended a talk by Ratan Tata, who recalled that he kind of blurted out in an interview that they were looking to develop a $2,500 automobile. At the time, everybody thought he was crazy, including many people in his own automobile company. So

he brought together a group of engineers for a significant period of time to continually reengineer the car, bring down the cost by making sure that everything included was essential, introducing new materials, and doing a very thorough job of engineering. Eventually they reached his target, and the cars are now being sold. Mr. Tata pointed out that much of the world is still poor, and it is very important to focus our efforts on providing additional capabilities for those who cannot afford what have been considered low-cost devices. A lot of progress is being made.

New fields are also moving ahead rapidly. I just attended an advisory council meeting for the School of Engineering at the University of California, San Diego, where I heard a discussion about new academic areas. One was nanotechnology, which I suspect is true at many schools. The nanotechnology program at UC San Diego, which has been in existence for about three years, now has a faculty of 17 and offers both undergraduate and graduate degrees in the field. The program comprises a wide range of topics, including photonics, materials, biomedicine, mechanical devices and motors, and energy storage, that will increasingly impact our lives.

There was also an interesting talk on composite materials. We are all familiar with the increasing use of composite materials in commercial aircraft design to reduce weight, increase strength, reduce metal fatigue, and allow higher cabin pressurization. At UC San Diego, they are studying these materials, but also studying safety issues and monitoring devices. For example, when struck, composite materials may spring back and not show a dent, even though they may have sustained internal damage. One member of the council noted that remote piloted vehicles are already constructed largely of composites and that many lessons have been learned.

The UC program is also looking into bridges made of composite materials and, interestingly, substituting composite materials for steel in the blades of wind turbines. So engineering innovation may lead to a gain in market share for wind turbines.

Another area of discussion was the Institute of Engineering in Medicine, which is expanding cross-disciplinary activities in bioengineering, biology, and clinical medicine. When that institute was established, a keynote talk was given by one of our new members, **Dr. Roderic Pettigrew**, who described a wide range of possibilities for engineering in medicine.

In addition, interest in telemedicine by academia, industry, and clinicians is growing, again exploiting the increasing power of wireless devices. One focus is on the use of biologic sensors, some that contact the body directly (often referred to as smart bandages) and some that can sense from a short distance. Better monitoring could greatly improve health care and reduce costs.

Indeed, these are very exciting times for engineers, but we do have a major failure. That failure is our inability to communicate the excitement of engineering to young people growing up in this country. As a result, we continue to lose ground to other countries in K–12 and STEM education, do not attract enough domestic students to college engineering programs, and are not gaining support for putting greater emphasis on and more resources into STEM education.

Although resources and teacher training will continue to be limitations, I believe technology will support major transformations in education over the next ten years. Yesterday, at an NAE Council meeting, some optimistic and some pessimistic points of view were expressed about education. I tend to be more on the optimistic side, although there is clearly lots of work to be done. But we already have a number of positive examples.

Qualcomm, other business leaders, and my family have been involved in San Diego with a charter school system called High Tech High that includes primary, middle, and high schools. Students are chosen by lottery from each zip code in the district to ensure that there is a very good mix. These students are encouraged to prepare for college and to study STEM subjects.

High Tech High has been going on now for about 10 years and has been quite successful—almost 100 percent of the kids graduate and continue their education, mostly at four- and sometimes two-year colleges. In a recent survey, 34 percent of the High Tech kids in college were studying STEM subjects, much higher than the national average.

How is that achieved? The High Tech High program is project-oriented—it includes many projects that introduce engineering design concepts and require research to complete. Students work on these projects in small groups, learning to work together and to support one another. I believe these social connections are important. Students working as part of a group get to know each other well, and if one is having problems, the others tend

to help. I think this peer-to-peer support is one of the reasons the graduation rate is so high. Another factor is that teachers are selected competitively from a large group of applicants, and subject knowledge is a priority.

Another program with which Qualcomm has been involved, called Project K-nect, provides smart phones with Internet access for use at school and home to students in several high schools in North Carolina. This program has now been running successfully for several years. Comparing classes taught by the same teacher, one class with smart phones and one without, has revealed significant differences in performance in several different subjects. For example, 100 percent of the kids with smart phones pass the statewide algebra test compared with only 67 percent in the class without the phones.

How is this achieved? There are many reasons, including enhanced curricula, remote tutoring support, and the ability to more easily communicate with a teacher. But perhaps the most important is peer-to-peer communication and support—academic social networking. For example, if a student at home gets stuck on a homework problem in the evening and cannot get help from parents, he or she can go online and check with other students in the same class around the state, getting help to move ahead rather than delaying until morning. In addition, a student who comes up with a good solution to a problem can video him or herself and make it available to others. The ability to supplement in-class teaching with additional work outside of class, to share information, and support other students peer-to-peer is, I believe, a key factor in the success of the program.

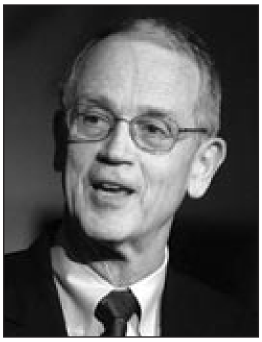
K-nect's success was recently recognized by the Federal Communication Commission (FCC), which set up a program called 2011 Pilot EDU, which stands for E-rate Deployed Ubiquitously. The program provides wireless interconnectivity 24/7 for poor schools. The FCC will evaluate the results at several schools to determine the effectiveness of this support and decide whether the program should continue and be expanded.

Although there are always downsides, for example, students distracted by other uses of the phones in the classroom, I believe technology-assisted education, combined with expanded use of electronic textbooks and materials, will transform and radically improve education in the coming decade.

Once again, I congratulate all of you in the incoming class. Best wishes for a productive membership in NAE, and I look forward to hearing more about your accomplishments.

Indeed, it is a very special time to be an engineer!

Technology and the Future of U.S. Competitiveness: Nightmares and Dreams—Remarks by NAE President Charles M. Vest



Charles M. Vest

In 2005, the National Academies responded to a call from a bipartisan group of senators and representatives for the top 10 actions by federal policymakers to enhance the U.S. science and technology enterprise so the United States can compete, prosper, and be secure in the global community of the 21st century. They also asked for a strategy, with several concrete steps, for implementing these actions.

Our response was *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*, a report by a committee of 21 distinguished leaders experienced in industry, academia, philanthropy, and government. The committee was chaired, with remarkable effectiveness, by **Norman Augustine**.

Gathering Storm stimulated a great deal of conversation in Washington and throughout the country. Together with work by the Council on Competitiveness, *Gathering Storm* led directly to passage of the America COMPETES Act of 2007, which authorizes implementation of most of our recommendations. The act was passed with remarkably large bipartisan votes and was signed by President Bush. Major components of the research budget we recommended have been

funded during both the Bush and Obama administrations, but this funding is "metastable" at best. It was added to once by a supplemental appropriation and is currently funded largely through the FY 2010 Stimulus Bill.

ARPA-E, a venturesome new energy research office, has been established as recommended by our committee and is funding a plethora of exciting, high-risk, high-benefit R&D in small companies and universities. Unfortunately, our highest priority recommendations, the ones that deal with K–12 science, technology, engineering, and math (STEM) education and the preparation of a 21st century teacher corps, have not yet been substantively addressed.

Today we face metastable research funding, insufficient action in K–12 education, and the expiration of the COMPETES Act at the end of this year. For these reasons, my colleagues, NAS President Ralph Cicerone, IOM President Harvey Fineberg, and I asked Norm Augustine and the available members of the original committee to saddle up again, scan the economic and technological horizons, and see how our nation's competitive position has changed in the five years since their report was issued.

On September 23, 2010, we released their unanimously approved report, *Rising Above the Gathering Storm Revisited: Rapidly Approaching Category 5*. The subtitle, *Rapidly Approaching Category 5*, says it all. The committee's overall conclusion is that "in spite of the efforts of both those in government and the private sector, the outlook for America to compete for quality jobs has further deteriorated over the past five years."

The members of the committee continue to believe that the critical underpinnings of a successful nation in today's global context are encapsulated in their recommendations, which are divided into four groups:

1. Increase America's talent pool by vastly improving K–12 science and mathematics education.
2. Sustain and strengthen the nation's traditional commitment to long-term basic research.
3. Make the United States the most attractive setting in which to study and perform research so that we can develop, recruit, and retain the best and brightest students, scientists, and engineers from the United States and throughout the world.
4. Ensure that the United States is the premier place in the world to innovate, invest in downstream activities such as manufacturing and marketing, and create high-paying jobs based on innovation.

So why did they conclude that our situation has deteriorated? I can only scratch the surface here, but let's start with the fact that we in the United States have always considered ourselves to be *Number One*. But here is a little dose of reality about where we actually rank today:

- # 6 in global innovation-based competitiveness, but #40 in rate of change over the last decade (ITIF, 2009)
- #11 among OECD countries in the fraction of 25 to 34 year olds who have graduated from high school (the older U.S. workforce ranks first among OECD populations of the same age) (OECD, 2009a, Chart A1.2)
- #16 in college completion rate and #20 in high school completion rate (OECD, 2009a, Chart A3.1)
- #22 in density of broadband Internet penetration (Dutta and Mia, 2010)
- #24 among 30 wealthy countries in life expectancy at birth (OECD, 2009b)
- #27 among developed nations in the proportion of college students receiving undergraduate degrees in science or engineering (OECD, 2009b, Table A3.5)
- #48 in quality of K–12 math and science education (World Economic Forum, 2009)
- #72 in density of mobile telephony subscriptions (Dutta and Mia, 2010)

This is not a pretty picture, and it cannot be wished away. As Bill Gates has said, "When I compare our high schools to what I see when I'm traveling abroad, I'm terrified for our workforce of tomorrow."

Successful entrepreneur Larry Bock says, "I find myself hiring talent for my companies abroad, not because I want to but because I can't find qualified engineers and scientists in America." By the way, Larry has taken the initiative by founding and driving the USA Science and Engineering Festival, a nationwide festival that will culminate in an expo on the National Mall here in Washington. NAE will have a great exhibit designed and implemented in partnership with Disney/Pixar to attract and inspire young people (see p. 77).

At our NAE Forum last year, NAE member and former DuPont CEO Chad Holliday said, "[Other

nations] have taken this recession [as an opportunity], not to talk about it, not to debate it, but to actually take steps … We must do exactly the same thing."

To drive the point home, I quote China's premier, Wen Jiabao, "The history of modernization is in essence a history of scientific and technological progress. Scientific discovery and technological inventions have brought about new civilizations, modern industries, and the rise and fall of nations . . . I firmly believe that science is the ultimate revolution."

Gulp.

Fact-based pessimism is in the air, and a national nightmare could be unfolding. But this does not have to happen. Even though it is the 11th hour, this nightmare need not materialize. Indeed, I do not believe it will. But we must get started now on a strategic agenda for the long haul.

The *Category 5* committee believes that implementing its recommendations is an essential foundation for such a national strategy. It is by no means the entire foundation, but it is the core of that foundation.

Three Pushback Questions

When I have engaged in discussions about *Rapidly Approaching Category 5* or about *Rising Above the Gathering Storm* or for that matter in any discussion about engineering and the future, three questions almost always arise:

1. Why should we invest in research when our discoveries and inventions will inevitably be snapped up and commercialized in other countries?
2. Why do we need more engineers or scientists when people in other parts of the world can do their work for a fraction of their salary or wage?
3. How can you request more federal expenditures for research and education in this time of financial hardship and growing deficits?

These indeed are vexing issues, because one can find a lot of data that support—or seem to support—these positions. The pool of talented, well educated, or well trained workers in other countries, especially in Asia, is growing at a dramatic rate. And the salaries and wages for professionals and skilled workers in China, India, Vietnam, and elsewhere *is* much lower than in the United States. We cannot control that. It is a direct result of rapid economic growth, supply and demand, market forces, stage of development, and also of history and culture.

These are daunting challenges for us and for the generation of Americans to follow. But inevitable? I can't see an inevitable devastating result here. Cause for profound worry? Absolutely yes. But I do not accept inevitability. Other countries cannot impose inevitability on us. It can only result from a loss of will or lack of logical response in our own country.

We all look back with great amusement at the prediction commonly attributed to patent commissioner Charles H. Duell in 1899 that "Everything that can be invented has been invented." Hopefully our grandchildren will look back someday and chuckle about those who at our turn of the century said, "Everything that can be invented in the United States and create domestic jobs has been. It's over."

America and Japan: Lessons Learned?

Perhaps we should pause and look back just 25 years and recall that in the 1980s, many serious people were convinced that Japan would dominate the world economy and that America would be crushed. That did not happen, and part of the reason it did not happen holds an important lesson for us today.

At the time, Japan had some major advantages. Its postwar generation worked incredibly hard in a very disciplined way. It had the advantages of building Greenfield factories, and—yes indeed—it had comparatively low labor costs. Its markets were either closed to us or were difficult for American companies to penetrate. It developed excellent engineering talent with a drive to excel, a deep attention to detail, and a respect for manufacturing. Japanese engineers set and accomplished bold goals for precision, performance, and miniaturization of consumer products.

The deep paranoia in the United States was based on a singular vision. Japanese students and visitors to the United States would take advantage of our open society by learning our technology and copying our innovations. They would then commercialize them and beat us economically—steal our technological crown jewels and run us over with them in the marketplace.

But what really happened? I contend that, in the end, the most important transfer of knowledge was not U.S. technology going to Japan but Japanese knowledge of manufacturing processes and quality transferred to the United States.

It was a painful period for us, but our consumer-manufacturing sector was forced to respond, and

it transformed itself. This transformation was hard, and it is still going on, but in the end, because of our own actions, the Apocalypse never happened. Today, Japan is indeed a prosperous nation, and the quality of life of its people has increased dramatically. Despite its subsequent economic stagnation, it is still the second or third largest economy in the world.

So nothing was inevitable. But U.S. companies had to understand the changing reality and adjust to it by transforming themselves. They had to rebalance their competition and cooperation with Japan. We competed fiercely, and by accepted standards the Japanese were not always fair competitors. But there was an odd sort of cooperation as we learned from them how to produce high-quality products with previously undreamed of specifications, throughputs, and cycle times.

By facing reality and acting, the United States was able to persevere, and indeed entered an unprecedented era of economic growth and wealth generation as our entrepreneurial spirit exploded and our past investments in basic research led to a vast array of new products and services. Among other things, we created the IT industry and launched the biotechnology industry.

Unfortunately, we lost that edge when we became overzealous about technology-based products that served very little real purpose. Market forces sorted that out, and the "tech bubble" burst, but we were left with some very important and, indeed, transformational companies and tools, *Google* being the prime example.

Responding to Category 5

In the last few years, a very different wave of economic damage has hit us. In my opinion, it had two fundamental causes. First, markets were entirely distorted as unfathomable amounts of capital came under the control of people and organizations whose work added little if any actual value to the economy. We forgot basic things, such as that the purpose of houses is to provide shelter and a decent quality of life for families and individuals and that the purpose of banks is to safeguard people's money and provide loans at reasonable rates to individuals and businesses for legitimate purposes. We also forgot that the very sophisticated computational tools and quantitative models we have produced for complex tasks, such as evaluating financial risk, can only be applied effectively by people who have an understanding of how they actually work and of the assumptions on which they are based.

On top of all of this, there has been devastating indifference toward the miserably inadequate way a very large fraction of our children are educated, blindness toward how dramatically the world as a whole and our place in it have changed, and refusal to face up to the results of our addiction to fossil fuels.

This is a very bleak analysis, and I believe we should be deeply worried. But my point in sharing my observations about the near-death experience of our manufacturing sector 25 years ago when Japan grew large on the world economic stage is that, once the truth sank in, we took the painful steps required to get back in the game. We analyzed, repositioned, persevered, and emerged stronger. We did it. In that case, the "we" was U.S. industry.

But this time around, more is required than change within companies. This time around we need a public awakening, establishment of political will, resetting of priorities, sacrifice for the future, and an alliance of governments, businesses, and citizens. This time we need truth telling, sensible investment, a rebirth of civility, and a cessation by both political and corporate leaders of pandering to our baser instincts.

Engineering, education, science, and technology are clearly critical to what has to be done. After all, this is the Knowledge Age. The United States cannot prosper based on low wages, geographic isolation, or military might. We can prosper only based on brainpower—properly prepared and properly applied brainpower.

This brings me back to the three questions frequently asked in response to the *Gathering Storm* and *Category 5* reports:

1. Why should we invest in research, when our discoveries and inventions will inevitably be snapped up and commercialized in other countries?

My answer is neither new nor original, but I believe it is correct. Robert Solow's well known, Nobel-Prize earning research taught us that the most important driver of productivity is technical progress driven by investment in research and new knowledge. Taking a very long view, Paul Romer's analysis of economic growth in the United Kingdom and the United States over a period of two centuries shows that it was only possible because of the continual development and advancement of technology.

Where do new technology and new knowledge come from? They come from research. And where do really transformative innovations come from? They come from

long time-horizon research—fundamental research and use-inspired basic research. Federal investment in U.S. university-based research brought us the computer, the laser, the enabling technologies of the GPS system, numerically controlled machines, the Internet, the deployment of the World Wide Web, the genetic revolution, and most of modern medicine. There are virtually no jobs in the United States today that are not directly enabled by one or more of these research-based innovations.

If we do not invest vigorously in basic research, an economic downslide is assured. If we do invest vigorously in basic research, we have a chance. By being first out of the box and increasing the probability of transformational breakthroughs, we can be first to produce and first to market. If we look clearly and holistically at our innovation system, we should be able to carve out job-producing space, especially at the high end.

I will be the first to admit that there are no guarantees and that tax policy, health care, patent protection, and other major factors must also be addressed. But no research, no chance.

2. Why do we need more engineers or scientists when people in other parts of the world will do the work for a fraction of the salary or wage?

I for one am not ready to fold up the tent and leave the field of competition. Dare I point out that salary and wage rates in every other category of employment are also lower outside the United States? So at heart, no matter how well informed or well intentioned, this perspective reflects a "*can't do*" mentality.

A month or two ago, it was reported that new manufacturing jobs were beginning to emerge in the United States, but workers with appropriate technical and quantitative skills are not available to fill them. A basic finding of the study of economic development is that the larger the number of technologically trained and creative people who come in contact with each other, the higher the probability of innovation. Floyd Kvamme, a major venture capitalist in the development of Silicon Valley, defines venture capital as "the search for good engineers."

Across Asia today, 21 percent of university graduates are engineers. Across Europe, 12.5 percent of university graduates are engineers. In the United States the number is 4.5 percent. So why hasn't the United States already been steamrollered? The answer is clear. We have addressed the engineering gap by attracting remarkably talented people from around the world to study in the United States and have been fortunate that many have stayed and become leaders in industry, academia, and entrepreneurship.

Large numbers of such individuals still aspire to stay and contribute to the United States, but our visa policies are making that path increasingly difficult to follow. We must fix this *post haste*. Furthermore, this gravy train is slowing down. Larger numbers of engineers and entrepreneurs are returning to China, India, and elsewhere. Vivek Wadhwa's surveys have shown that the primary reasons they are returning are that their professional careers or the companies they wish to found can be built much faster back in Asia or South Asia.

I believe we need more U.S. engineers to create and lead the companies, products, services, and processes of the future. Despite the horrendous global headwinds that are rapidly approaching Category 5, there is still value in locating companies and manufacturing facilities where the smart and innovative engineers are.

3. How can you request more federal expenditures for research and education in this time of financial hardship and growing deficits?

Our colleague and peerless leader Norm Augustine answers this question based on his experience as an aeronautical engineer and business executive: "When it becomes necessary to reduce the weight of an airplane, you don't accomplish it by throwing off the engine."

Why do we react to this by laughing? It is because we immediately and nervously recognize that this is a clear truth and an accurate analogy. It is not flippant. It makes a valid and essential point.

Nightmare or Dream?

It's time we regained our optimism and our "can do" spirit so we can remain a great nation and meet the challenges of our time. The way to accomplish this is to reconnect what we do with what we dream. We need a country with more people dreaming about what's possible, where young people—no, all people—are inspired to imagine a better world and help make it a reality.

That was once the American way, but now we are wandering around seemingly aimlessly. What happened to the charge-ahead spirit that led to the success of our "moon shot" challenge? Today we carry far more computing power in our pockets than was on an entire Apollo spacecraft,

but we have a lot less public passion for engineering dreams into reality. I don't mean mindless TV "reality," I mean real reality—improving the lives of real people and creating real jobs.

In the last century, big-thinking engineers brought us automobiles, airplanes, electrification, clean water, computers, refrigeration, radio, television, medical imaging, lasers, the Internet, and the Web. They transformed our world. Those engineers were mostly young, and they were empowered by education and funded by government, industry, and venture capital to create new technologies, hire people to produce them, and move them into the marketplace. That was the heart of the *real economy*.

Yes, some of those technologies also left a legacy of problems we now must deal with, such as cybercrime, the specter of nuclear war, and a national addiction to fossil fuels. But if we can inspire, educate, and fund them, a new generation of engineers can be at the heart of solving these problems and of making dreams of a better world become the new reality.

Our 14 *NAE Grand Challenges for Engineering* address energy, water, climate and sustainability; improving the delivery of health care; increasing security against both natural and human threats; and expanding human capabilities and joy. Some of these challenges must be met to sustain human life on Earth. All of them, if met, would improve the quality of life on Earth.

Working to address the Grand Challenges should be made as appealing to people—especially young people—as excelling in sports or acting. The United States used to be full of people who believed in endless possibility, but pessimism is now holding us back. As my favorite philosopher, Pogo Possum, said, "We are surrounded by insurmountable opportunities."

I would like to challenge Congress to reauthorize and fund the America COMPETES Act to help propel us back to 21st century innovation by training and rewarding competent and inspiring teachers; once again attracting the best and brightest minds from America and the world to study science, engineering, and mathematics; and supporting the fundamental research necessary to power our economy by creating real value.

And I'd like to challenge all of us to stop shortchanging our children by failing to provide them with a world-class education that both inspires them to dream big dreams and empowers them to make those dreams real. I am optimistic. Puzzles, problems, questions, challenges are what inspire young people. Want to see a kid crave science? Give her a cause. Let her know she can use science to change the world.

Dozens of universities across the United States now provide that opportunity through the NAE Grand Challenges Scholars Program, which will prepare students to be the generation that can tackle the big issues facing society. These scholars will build on a core of technical education, but will also be able to join forces with colleagues from humanities, management, political science, and law to meet the challenges of our time. Because of such initiatives, there are still Eureka! moments ahead of us. But we must draw young people in and excite and prepare them.

Re-defining who we see as heroes, perhaps with the help of the entertainment industry, is part of the answer. Americans need to watch, read, support, and demand what is important as well as what is entertaining. Artists can help open our minds, and athletes can show us the power of focused excellence, but we will need engineers to turn visions into reality. It is time to change the national conversation and the national agenda, because dreams need doing. Nightmares don't!

So, what will it be, nightmare or dream? It is our choice to make.

References

Dutta, S., and I. Mia. 2010. Global Information Technology Report 2009–2010: ICT for Sustainability. Geneva, Switzerland: World Economic Forum.

ITIF (Information Technology and Innovation Foundation). 2009. The Atlantic Century: Benchmarking EU and US Innovation and Competitiveness. February. Available online at *http://www.itif.org/publications/atlantic-century-benchmarking-eu-and-us-innovation-and-competitiveness*.

OECD (Organization for Economic Cooperation and Development). 2009a. Education at a Glance, OECD Indicators, 2009. Available online at *http://www.oecd.org/dataoecd/ 41/25/43636332.pdf*.

OECD. 2009b. Health at a Glance. Available online at *http://www.oecd.org/health/ healthataglance*.

World Economic Forum. 2009. The Global Competitiveness Report 2009–2010. Available online at *http://gcr.weforum.org/gcr09/*.

2010 Founders Award Acceptance Remarks by Robert Langer

NAE president Charles M. Vest, Founders Award recipient Robert Langer, and Rod C. Alferness, 2010 NAE Awards Committee Chair and chief scientist of Bell Labs.

The 2010 Founders Award was presented to **Robert Langer**, David H. Koch Institute Professor, Massachusetts Institute of Technology, "for the invention, development, and commercialization of methods and materials for drug delivery and tissue engineering, mentoring of scientists, and the promotion of the nation's health." His acceptance remarks follow.

I am delighted to be here to accept this wonderful award. In preparing what I would say today, I thought it would be useful for me to read the remarks of previous winners such as **Bob Nerem** and **Shu Chien**. They spoke about their lives and careers, which I found fascinating and inspiring. I'm not sure my life story is that interesting, but I feel incredibly fortunate that I am here today, particularly given my struggles in my early career. I'd like to speak to you today about that time in my life.

When I finished graduate school in chemical engineering, I didn't know what I wanted do career-wise. I graduated in the 1970s, when, just like a few years ago, there was a gas shortage. But it was even worse then. Not only did the price of gas go way, way up, but you also had to wait in line at the gas station for hours to fill up your car. Consequently, if you were a chemical engineer, you got a lot of job offers.

In fact, nearly all of my classmates in the 1970s joined oil companies, which had many openings, and that's where all the high-paying jobs in chemical engineering were at that time. I actually got 20 job offers from oil and chemical companies—four from Exxon alone. I also got offers from Shell and others. Maybe the only oil company I didn't get an offer from was British Petroleum.

One job interview made quite an impression on me. The interview was at Exxon in Baton Rouge, and one of the engineers there said that if I could increase the yield of a particular petrochemical by 0.01 percent, that would be wonderful. He said it would be worth a fortune. I remember flying home to Boston that night, thinking to myself that I really didn't want to do that.

What did I want to do? I had a dream of using my background to improve people's lives. I had spent a lot of my time as a graduate student establishing a school for poor high school kids and developing new chemistry and math curricula. One day, I saw an advertisement for an assistant professor at City College of New York to develop a chemistry curriculum. So I wrote them a letter, but they didn't write me back. However, I liked the idea, so I found all the ads I could for assistant professor positions to develop chemistry curricula. I wrote to all of them, but very few wrote me back, and the few that did said they were not interested.

Another way I thought I could help people was through health-related research. So I applied to a lot of hospitals and medical schools, but none of them wrote me back either. Then one day, someone in my lab said I should write to Dr. Judah Folkman at Harvard, because "he sometimes hires unusual people."

So I took what, at the time, seemed to all engineers like a huge risk and began doing postdoctoral work in a hospital. It might seem more common today, but at that time few if any chemical engineers had ever done postgraduate work in a surgery lab. The projects I began working on involved two related problems: (1) trying to discover the first substance that could stop cancer blood

vessels (and thus stop cancer) and (2) developing plastics that might be able to slowly release these and other large molecular-weight substances over a very long time in the body.

Before I tackled the second problem, *no one* had been able to develop a way to slowly release these kinds of substances over a long period of time. In fact, scientists and the literature said this was *impossible*. In fact, maybe the only thing I had going for me was that I just didn't know that. So I tried it anyhow.

After two years, I found 200 different ways to get it to not work. But finally, I discovered that I could modify certain types of plastics and use them to slowly release those molecules. We used this to discover the first substances that stopped cancer blood vessels and helped stop cancer.

About two years after I started working on this problem, I was asked to give a talk on my work to a very distinguished audience of polymer chemists and engineers in Michigan. I had never given a big talk before. (PAUSE) Well actually, in 8th grade I had to give a minute-and-a-half speech. So the night before my 8th grade talk, I rehearsed it for four straight hours in my parents' bedroom in front of a mirror. The next day, I started to give the talk, but after one minute of speaking, I couldn't remember the next word, and I froze. Eventually, the teacher told me to sit down and gave me a very poor grade. I think it was an F.

So when the Michigan talk came along many years later, I was very nervous. I stopped working two weeks in advance of it, and I practiced my talk over and over into a tape recorder until, finally, the day came when I was supposed to give it. I got up and gave that talk, and I actually was pretty pleased by the end of it. I hadn't forgotten too much of what I'd intended to say, and I didn't stammer or stutter too much. And I thought that when I was done with that talk all these much older, distinguished engineers in the audience, being nice people, would want to encourage me, this young guy.

Intstead, a number of people gathered around me and said, "We don't believe anything you've just said. We know that you can't get these molecules that you're talking about through these plastics." It wasn't until several years later, when other people began repeating what we had done, that the question shifted to "How could this possibly happen?" In fact, I spent a good part of my early career at MIT trying to understanding how plastic drug-delivery systems functioned and making them useful for different medical applications.

Shortly after that talk, I also tried to obtain funding to support my research by writing a number of grants. The first nine were turned down. I remember I wrote one grant to the National Institutes of Health for some cancer research. When I got the reviews back, they were very, very negative. They not only turned me down, but they said, "How could Dr. Langer do this? He's an engineer. He doesn't know anything about biology, and he knows even less about cancer?"

When I finished my postdoctoral work, I applied for faculty positions in a number of chemical engineering departments, but I had trouble getting a faculty job because people at that time felt that what I was doing wasn't engineering. They thought it was more biology. So I ended up joining what was then the Nutrition and Food Science Department at MIT. But in that department, the year after I got the position, the chairman who had hired me left, and a number of senior faculty decided to give me advice. They told me I should start looking for another job.

So there I was, my grants being turned down, people not believing in my research, and having little hope of even keeping my job—and it was the lowest-level academic position one could have. I was fortunate, however, that within a year or two, scientists and engineers in the pharmaceutical industry started using some of the principles, and even some of the inventions I had made, and things began to turn around. Eventually, I did get promoted. More important, I like to hope and think that the discoveries we've made both in cancer and drug delivery have helped improve the lives and health of millions of people.

I'm very honored that I'm here today to receive the Founders Award. When I look back at how my career started, I feel tremendously fortunate to receive this wonderful prize. Thank you so very much.

2010 Arthur M. Bueche Award Acceptance Remarks by Anita K. Jones

Charles M. Vest, Bueche Award recipient Anita K. Jones, and Rod C. Alferness.

The 2010 Arthur M. Bueche Award was presented to Anita K. Jones, University Professor Emerita, University of Virginia, "for leadership in the development of U.S. science and technology policy and the development of technologies for national security, including technical contributions to high performance computing and cybersecurity." Her acceptance remarks follow.

Mr. Chairman, Mr. President, ladies and gentlemen, I am deeply honored to receive the Arthur M. Bueche Award from NAE. It comes from an organization I revere, and it honors me greatly by association with previous winners.

I grew up a girl in Texas. I chose computer science as my field of study and later joined the faculty of Carnegie Mellon University, one of the premier computer science departments in the world. Yet, after a few years in teaching and research, I was feeling a little bored.

An NAE member, **Keith Uncapher**, had made himself my mentor and was watching over my career progression. I told him about feeling insufficiently stimulated professionally. Keith called back the next day and said, "I thought about it, and you should join the Air Force." After the shock of incredulity passed, I found that he had arranged for me to become a member of the Air Force Science Advisory Board, a group of about 30 individuals who advise the U.S. Air Force leadership.

Graham Greene wrote, "Once in a while a door opens and the future walks in." For me, the future, had, indeed, walked right in. For the next several years—still wet behind the ears and 10 to 20 years younger than most of the other board members—I sat at conference tables with people with extraordinary minds, and I watched and learned how they approached problems of great magnitude. I was privileged to work with people like Glenn Kent, **Gene Covert**, **Al Flax**, and **Natalie Crawford**—a number of them engineers—and I learned that I could contribute to solving major problems.

Sometime during this life-altering education with the Air Force Scientific Advisory Board, I came to know that I was passionately devoted to advancing the U.S. innovation enterprise, especially research activities that routinely create new knowledge and train the experts who will be needed tomorrow. Over the years, I have taken on many assignments in this arena, and I learn something new with each one. I would like to share some of those lessons with you:

1. I learned that the best way for our federal government to nurture research is to have mission agencies fund basic and applied research in areas that promise to serve their needs. I saw this most clearly when I oversaw the U.S. Department of Defense (DOD) science and technology program. After World War II, President Roosevelt had asked Vannevar Bush how to ensure that scientists and engineers would be available to the nation in a future crisis as they had been during that war. In Bush's report, *Science, the Endless Frontier*, he proposed that there be a single federal agency to fund curiosity-driven science and engineering in universities (that agency is the National Science

Foundation [NSF]). In addition, several U.S. mission agencies such as DoD have—decade in and decade out—sustained merit-based research, choosing topics guided by their missions. With some frequency, this mission-driven research has yielded results that are different and sometimes better than those generated by curiosity-driven research. In that respect, Bush was wrong. History shows that a combination of curiosity-driven research and mission-motivated research yields the best results.

2. Serving on the National Science Board—the governing board of NSF—reinforced my belief that it is useful to have an agency dedicated to funding curiosity-driven research. Specifically, NSF ensures stable funding for students as well as for the construction and operation of large scientific instruments, neither of which may have a high priority for program officers in mission agencies. The two types of funding guided by different considerations are important to maintaining a healthy research and education enterprise.

3. The "wall of classification" around intelligence agencies is impenetrable to small entrepreneurial companies that may be developing exactly the new technologies those agencies need. I learned that creative government action can broach that wall. I am a trustee for the not-for-profit organization called InQTel, which was created and funded by insightful leaders at the Central Intelligence Agency. InQTel has invested in more than 100 high-tech companies, some of whose products have been piloted and adopted by multiple intelligence agencies. This would not have been possible without InQTel acting as intermediary.

4. I learned that even a small country can improve its ability to compete in the burgeoning knowledge economy. I learned this by serving as a founding trustee of Science Foundation Ireland (SFI). Designed to be an Irish clone of NSF, SFI funds merit-based research in Irish universities and has lured émigrés back to Ireland who had become distinguished researchers in other parts of the world. SFI has genuinely advanced Ireland's ability to compete in the knowledge economy.

5. Today, I am studying whether a single state, Arizona, can improve its ability to create new jobs by funding merit-based university and industrial research in the state. I am a founding trustee for Science Foundation Arizona, essentially another NSF clone. Time will tell if this initiative succeeds.

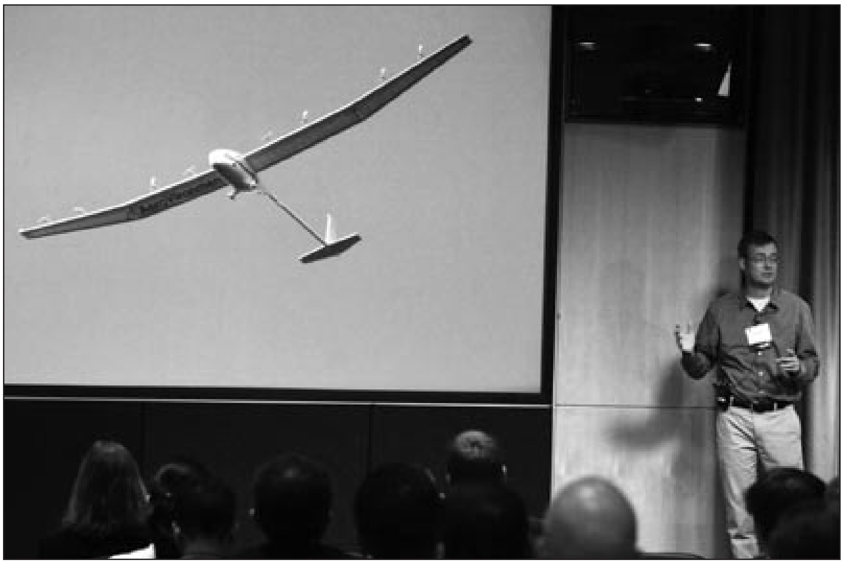
So, I have been privileged to be a participant in many different activities that have helped make the U.S. innovation enterprise vibrant and adaptive. With all its chaos, its many dimensions, and (thankfully) its lack of central planning, our innovation enterprise remains resilient. But other nations are attempting to compete and lead in technological innovation.

I participated in the National Academies committee that reported—five years ago—that a *Gathering Storm* threatens the U.S. quality of life. One of the issues we raised was that the United States had not been paying enough attention to core problems such as quality of teaching and research investment. Just this month our *Gathering Storm* committee issued a second report asserting that, as a result of this neglect, the storm is now approaching "Category 5 status."

I have lived through the decades during which the participation of women scientists and engineers has increased dramatically. Having chosen computer science as a professional field, I have watched advancements in information technology power the strongest economic force in the world over the past half-century. It is a privilege to play even a small part in that and to work on problems that really, really matter.

Thank you for this honor.

2010 U.S. Frontiers of Engineering Symposium



As co-chair for the session on Autonomous Aerospace Systems, Jack Langelaan (Pennsylvania State University) helped introduce the topic.

On September 23–25, NAE welcomed 109 mid-career engineers to the 2010 U.S. Frontiers of Engineering (US FOE) Symposium hosted by IBM at the IBM Learning Center in Armonk, New York. NAE member **Andrew M. Weiner**, Scifres Family Distinguished Professor of Electrical and Computer Engineering at Purdue University, chaired the organizing committee and the symposium. The seven articles in this issue of *The Bridge* are based on presentations at the symposium. In addition, these articles and summaries of the other presentations will be published in February 2011 in the annual FOE volume.

Talks at the symposium covered topics in cloud computing, engineering and music, autonomous aerospace systems, and engineering inspired by biology. Talks in the cloud computing session focused on how this disruptive technology is changing the way users design, develop, deploy, use, and disseminate applications and data. Following an overview presentation on the potential of cloud computing, talks focused on the challenges of providing transparent interfaces to users, storing massive amounts of data in the cloud and achieving resiliency, the energy-consumption and environmental ramifications of cloud computing, and architecture and software designs that can reduce the carbon footprint of the supporting infrastructure.

The underlying premise of the session on engineering and music was that technology has strongly influenced music since the first musical instruments were developed and that it continues to shape the music industry through personalized music search and retrieval, fan-generated remixes and "mash-ups," and advanced tools for music creation. Presentations in this session covered advances in very-large-scale music-information retrieval; how people outside the engineering community are using technology to create music; the use of laptop computers, which now have sufficient processing power to perform complex synthesis and audio-transformation operations, in collaborative live performance; and the relationship between music and mathematics—specifically how mathematics is being use to analyze and understand musical structure and how mathematical representations are being incorporated into visualizations for live performance.

The third session, on autonomous aerospace systems, focused on a number of aeronautical and space autonomous systems, from single vehicles to teams of robots to applications in the National Air Transportation System. Speakers covered techniques for enabling "intelligence" in autonomous systems through probabilistic models of the environment and the integration of human operators into the control/planning loop; challenges for automation posed by NASA's current and future space missions; the importance of health awareness in systems of multiple autonomous vehicles; and the role of automation and autonomy in the deployment of the next-generation air transportation system.

The symposium concluded with a session on engineering inspired by biology—innovations in contemporary engineering from bio-inspired, biomimetic, and bio-derived technologies. Talks focused on engineering challenges in the analysis of genetic variation, gene expression, and function, which could lead to faster screening and detection of diseases and therapeutics based

on an individual's genetic predisposition; engineering biomimetic peptides for targeted drug delivery; and using biomolecules as motor-powered devices within systems.

On the first afternoon of the meeting, participants broke into small groups for "get-acquainted" sessions where each of them presented and answered questions about a slide describing their research or technical work. This event gave them an opportunity to get to know more about each other relatively early in the program. On the second afternoon, IBM conducted tours of the IBM T.J. Watson Lab in Yorktown Heights and the IBM Industry Solutions Lab (ISL) in Hawthorne. The tour of the Watson Lab included the Blue Gene supercomputer, efforts to synthesize grapheme and fabricate high-speed transistors, ultra-high concentrated photovoltaic solar systems, the IBM Speech-to-Speech Translation System, and the Translingual Automatic Language Exploration System. At ISL, participants learned about research in health care analytics, cybersecurity, and mobile applications to support collaboration.

The dinner speaker this year, Dr. Bernard Meyerson, IBM Fellow and vice president for innovation, delivered a compelling talk entitled "Radical Innovation to Create a Smarter Planet."

Breaks provide an opportunity for networking among participants. Pictured from left to right: Y.Y. Zhou (University of California, San Diego), Efie Kokkoli (University of Minnesota), Ana Arias (PARC), Shriram Ramanathan (Harvard University), and Stergios Roumeliotis (University of Minnesota).

Planning is under way for the 2011 symposium, which will again be chaired by Andrew Weiner. The event will be held September 19–21, 2011, at Google headquarters in Mountain View, California. The topics will be additive manufacturing, engineering sustainable buildings, neuroprosthetics, and semantic processing.

Funding for the 2010 U.S. FOE Symposium was provided by IBM, The Grainger Foundation, Air Force Office of Scientific Research, Defense Advanced Research Projects Agency, Department of Defense (ODDR&E), National Science Foundation, Microsoft Research, and Cummins Inc.

For more information about the FOE symposium series or to nominate an outstanding engineer to participate in a future FOE meeting, contact Janet Hunziker at the NAE Program Office at (202) 334-1571 or by e-mail at *jhunziker@nae.edu*.

First E.U.-U.S. Frontiers of Engineering Symposium

Participants at the first E.U.-U.S. Frontiers of Engineering Symposium on the lawn of Jesus College, Cambridge.

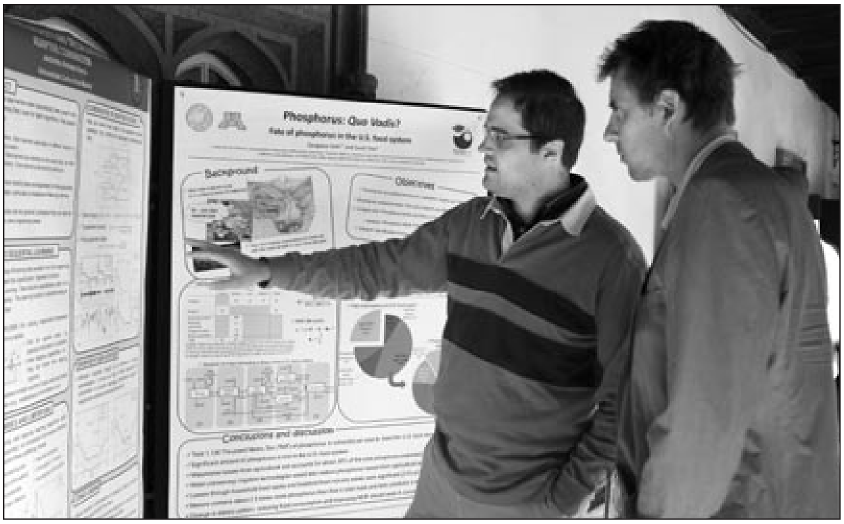
On September 1–3, NAE added a fifth bilateral meeting to its Frontiers of Engineering portfolio with the inaugural E.U.-U.S. Frontiers of Engineering Symposium (EU-US FOE) at Jesus College in Cambridge, U.K. The European Council of Applied Sciences and Engineering (Euro-CASE) co-sponsored the event, and organizational assistance was provided by the Royal Academy of Engineering. Co-chairs were NAE member **Sergio Verdu**, Eugene Higgins Professor of Electrical Engineering at Princeton University, and Dr. Richard Williams, Professor of Mineral and Process Engineering at the University of Leeds.

In the tradition of bilateral FOE symposia, the first EU-US FOE brought together approximately 60 engineers, ages 30 to 45, from U.S. and European universities, companies, and government laboratories to learn about leading-edge developments in four fields of engineering—bio-inspired engineering, signal processing, augmented reality, and materials ecology. Eleven E.U. countries were represented: Austria, Croatia, the Czech Republic, Denmark, France, Germany, the Netherlands, Spain, Sweden, Switzerland, and the United Kingdom.

Topics addressed in the session on bio-inspired engineering session focused on translations of biological principles to solve multidisciplinary technological challenges. Talks covered (1) improving our understanding of the physical principles of insect flight to shed light on efficient locomotion in fluids; (2) bio-concrete, or bacteria-based self-healing concrete; (3) investigation of how animals process sensory information to control locomotion, which has led to insights into designs for sensor-based robot-control systems; and (4) research in self-assembling nanoparticles for oncological treatments.

Signal processing, the subject of the second session, is a highly interdisciplinary field that can advance our capability to acquire, extract, and enhance information from sensors and other transducers. The unifying theme of this session was the concept of a signal model, a mathematical construct that makes it possible to distinguish between useful information and noise, for example. The first speaker discussed how diversity and specialization principles that have been successfully used in machine learning can also be used to improve the performance of classical digital signal-processing techniques. The second speaker described recent research on mitigating interference that threatens to clog wireless communication systems. The next talk was on graphical models, which combine concepts from graph theory with statistical modeling to enable new ways of finding patterns and explanations in a wide range of data types and applications. The subject of the fourth talk was compressive sensing—how models for compressing signals and images (e.g., MP3, JPEG, and MPEG) can be generalized to solve massive pattern-recognition problems in recommender systems (e.g., systems used by Amazon and Netflix).

The topic for the third session, augmented reality (AR), involves



Poster sessions provide an opportunity for participants to learn about each other's research and technical work. Poster 2: Jeronimo Arenas-Garcia (University Carlos III of Madrid) and Henk Jonkers (Delft University of Technology) discuss a poster.

improving our understanding of and interaction with the surrounding world by superimposing spatially aligned 3D graphics directly over a user's view of the real world. The topics of these talks were: displaying information using, for example, projector-based and head-worn displays; visual-tracking technologies to ensure that virtual graphics are drawn in the right place with minimal jitter or latency; the development of AR devices that will be as ubiquitous and affordable as mobile phones; and research in hybrid user interfaces and interaction techniques for unobtrusive AR.

Materials ecology, a synthesis of industrial ecology and materials life-cycle management, was the topic for the fourth and final session. The first speaker provided an overview of the topic, emphasizing its importance in responding to pressures from changes in population, technology, and consumption. The second speaker described the role of energy materials in different value chains and how changes in these materials influence industrial energy processes and the resulting ecological footprint. The third speaker posited that steel is a sustainable material; he described integration in the modern steel industry and highlighted the advantages of recycling steel and how the by-products of steel production are being used in other industries. The session concluded with a presentation on sustainable life-cycle management for one product—the automobile. The speaker identified challenges in terms of input constraints (e.g., materials input) and output constraints (e.g., emissions).

The two-and-one-half day symposium included a poster session on the first afternoon that was both an ice breaker and an opportunity for all participants to share information about their research and technical work. On the second afternoon, participants were given a walking tour of Cambridge that ended with punting on the River Cam. A dinner speech on the first evening was given by Sean Blanchflower, head of R&D at Autonomy, a company that resulted from research conducted at the University of Cambridge; Autonomy is now the second largest pure software company in Europe.

Financial support for the symposium was provided by The Grainger Foundation, which also provided core support that enabled us to initiate this new bilateral FOE program. Additional sponsors included numerous European organizations—Autonomy, ARM, Nanofactory, Fundación Pro Rebus Academiae, Real Academia de Ingenieria, OHL, Universidad Carlos II de Madrid—and the U.S. National Science Foundation.

The next EU-US FOE symposium will be held November 3–5, 2011, in the United States. **Sergio Verdu** and Richard Williams will continue to serve as symposium co-chairs.

For information about any FOE symposium series or to nominate an outstanding engineer to participate in a future Frontiers meeting, contact Janet Hunziker at the NAE Program Office at (202) 334-1571 or by e-mail at *jhunziker@nae.edu*.

First Annual USA Science and Engineering Expo

The TRON Lightcycle movie prop wowed visitors as they entered the NAE exhibit.

Approximately 500,000 people visited the National Mall in Washington, D.C., for the first USA Science and Engineering Festival and Expo. Held on October 23 and 24, the Expo was the culminating event of two weeks of activities and programs around the country to stimulate young people's interest in science, technology, engineering, and math. Exhibits by units of the National Academies featured a "distracted driver" simulator, hands-on games, and talks by Nobel Prize winners John Mather and Peter Agre, as well as many other speakers throughout the weekend.

Approximately 1,400 visitors—mostly families—came through the National Academies' tent ("Because Dreams Need Doing"). In one area, NAE and Walt Disney Studios teamed up to create an interactive exhibit that blended themes and images from the upcoming motion picture *TRON: Legacy* with NAE's Grand Challenges for Engineering. Visitors were treated to a demonstration of how 3-D scanning brings the real and virtual worlds closer together, tried their hands at virtual-reality "brain surgery" on a computer-generated replica of a real brain, and journeyed into the *TRON: Legacy* digital grid through a 3-D light painting activity created especially for the Expo.

NAE also coordinated three stage presentations. Lanny Smoot, a Disney Imagineer, gave a multimedia presentation on "The Top 10 Reasons You Might Be an Engineer." Next, IBM's "Watson," a Jeopardy-playing supercomputer, challenged the audience. The final presentation was a live demonstration by VICON Motion Systems of the motion-capture technology used in *TRON: Legacy*. In addition, NAE member and Draper Prize winner **Vint Cerf** was interviewed on stage by Joe Palca of National Public Radio.

NAE President Charles Vest and family enjoyed the 3-D scanning exhibit.

Mirzayan Technology and Policy Fellows

Chelsea Martinez

Lee Pearson

Chelsea Martinez is at the ABD (all but the dissertation) stage in her pursuit of a Ph.D. in bioorganic chemistry from the University of Texas (UT) at Austin, where she synthesizes tiny bio-inspired aromatic molecules and studies the weak noncovalent attractive forces that determine their 3-D shape in water. Chelsea graduated from Oberlin College in 2002 with majors in biochemistry and chemistry and then taught science at a Virginia boarding school and math at a California public high school before returning to the lab. As head teaching assistant for the most popular 500-person organic chemistry course at UT, she dressed up as Little Red Riding Hood, a Jimmy Buffet Parrothead, and a 5'6" electromagnet in the name of making premeds love chemistry. She also hosts a science news talk show on the UT student-run radio station.

Chelsea spent one summer interning at the *Los Angeles Times*

health desk and another interviewing physics Nobel Laureates and graduate students in Lindau, Germany, on a travel grant from the National Association of Science Writers. She hopes to build a career communicating science to nonscientists in a school, on the radio, or in some other venue. She spends her time between work and sleep reading maximalist fiction, swimming, defending the merits and beauty of the city of Los Angeles, and making quilts for friends who are hitting milestones. At NAE, Chelsea worked on the Program on Technological Literacy helping to create and review content for a new website being developed as part of an NSF-funded project to improve engineering messaging.

Lee Pearson recently completed an M.Phil. in Engineering for Sustainable Development at the University of Cambridge on a Marshall Scholarship. For his dissertation, he worked with the economic regulator of Scottish Water to analyze potential reductions in water demand in the residential customer base. Last year, he developed a linear programming model to optimize national water footprints through trade while working on an M.Sc. in ecological economics at the University of Edinburgh. Lee received his B.S.E. in civil engineering from Duke University. In December 2010, he will begin his Ph.D. studies in the Centre for Environmental Policy at Imperial College London.

Lee has eclectic interests. He has published articles on oil pipeline dynamics and the results of research on an NSF-sponsored program at Dalian University of Technology in China and conducted fungicide research with Bayer CropScience in Germany. As president of the Duke chapter of Engineers Without Borders (EWB-USA), he directed development projects in Peru and Uganda. While in England, he volunteered as a project manager for EWB-UK initiating placements for British students with a rural Ugandan NGO. Lee is also passionate about nature; he greatly enjoys camping, whitewater rafting, and hiking.

At NAE, Lee worked with Proctor Reid on the Roundtable on Technology, Science, and Peacebuilding, a new joint activity by NAE and the U.S. Institute of Peace.

NAE Members Testify before Congress on *Gathering Storm*

Four NAE members testified before the U.S. House of Representatives Committee on Science and Technology at a hearing on September 29. All four were members of the authoring committee of the influential 2005 report, *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*, as well as the follow-up report, *Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5*, which was requested by the presidents of the National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, and was released to Congress on September 23.

The focus of the testimony was on the urgent need for investments in science and education to improve the competitive position of the United States in the global economy. Among the committee's conclusions were that America's status in the competition for quality jobs in the global economy has deteriorated significantly in the last five years and that to keep it from slipping further, we will need a sustained investment in education and basic research to encourage and support innovation.

The four NAE members who testified were: **Norman R. Augustine**, former chairman and CEO, Lockheed Martin Corp., and chair of the study committee for both reports; **Craig R. Barrett**, former chairman and CEO, Intel Corp.; **Charles O. Holliday Jr.**, current chairman of Bank of America and former chairman and CEO, DuPont; and **C.D. (Dan) Mote Jr.**, President Emeritus and Glenn L. Martin Institute Professor of Engineering, University of Maryland.

2010 *EngineerGirl!* Essay Contest

For the past decade, NAE's *EngineerGirl!* website (*www.engineergirl.org*) has been a crucial resource for students considering engineering careers, as well as for their parents and teachers. This unique resource, which features more than 100 profiles of women working in engineering positions and a question-and-answer section where students can ask questions of practicing engineers, offers girls a glimpse into the world of engineering practice.

One of the most popular features on the site is the annual *EngineerGirl!* essay contest. Designed to excite students about the contributions of engineering to society and the betterment of the human condition, the contest provides an opportunity for them to put their knowledge and creativity to work in a national competition for cash prizes.

The topic for this year is Engineering and Human Service—Relief from a Disaster. As more and more disasters, both natural and man-made, make headline news, we are asking students to consider how engineers might help. Students in three age categories (elementary, middle, and high school) may submit essays describing something designed for use in a post-disaster situation. The essays must include descriptions of the design of the item and the engineer's role in designing it. The contest deadline is March 1, 2011. For more information, please see the contest webpage (*http://www.engineergirl.org/CMS/Contest/2011.aspx*).

Calendar of Upcoming Events

2011

January 1–31	Election of New NAE Members
January 3–April 1	Call for NAE Awards Nominations
February 4	Membership Policy Committee Meeting Irvine, California
February 7	The National Academies Corporation Board and Advisory Board Meetings Irvine, California
February 8–10	NAE Council Meeting Irvine, California
February 10	NAE National Meeting Irvine, California
February 20–26	National Engineers Week
February 22	NAE Awards Forum and Awards Dinner/Ceremony Union Station, Washington, D.C.
February 24–25	The Role of Engineering in STEM Education Workshop TBD, Washington, D.C.
March 1–31	Election of NAE Officers and Councillors
March 28–30	China-America Frontiers of Engineering Symposium San Diego, California
March 31	NAE Regional Meeting Berkeley, California

All events are held in the Keck Center in Washington, D.C., unless otherwise noted.

In Memoriam

DAVID H. ARCHER, 82, retired consulting engineer, Westinghouse Electric Corporation, and adjunct professor, Carnegie Mellon University, died on June 24, 2010. Dr. Archer was elected to NAE in 1989 "for leadership in developing coal-based energy systems."

GAIL DE PLANQUE, 65, president, Strategy Matters Inc., died on September 8, 2010. Dr. de Planque was elected to NAE in 1995 "for leadership of the national nuclear programs and contributions to radiation protection devices and standards."

JAMES R. FAIR, 90, John J. McKetta Centennial Energy Chair Emeritus in Engineering, University of Texas, died on October 11, 2010. Dr. Fair was elected to NAE in 1974 "for contributions to mass transfer technology and computer simulation of chemical processes."

JOSEPH G. GAVIN JR., 90, retired president, Grumman Corporation, died on October 30, 2010. Mr. Gavin was elected to NAE in 1974 "for leadership in the design and production of the Apollo Lunar Module."

ROBERT W. GUNDLACH, 83, retired senior research fellow, Xerox Corporation, died on August 18, 2010. Mr. Gundlach was elected to NAE in 1994 "for contributions to the development of xerographic copying and printing, including manifold inventions."

FREDERICK JELINEK, 77, Julian S. Smith Professor, Johns Hopkins University, died on September 14, 2010. Dr. Jelinek was elected to NAE in 2006 "for contributions to statistical language processing with applications to automatic speech recognition."

JAMES C. KECK, 86, Ford Professor of Engineering Emeritus and senior lecturer, Massachusetts Institute of Technology, died on August 9, 2010. Dr. Keck was elected to NAE in 2002 "for developing innovative, widely used new concepts for modeling coupled chemical and physical phenomena in engine combustion and high-temperature flows."

RUSTUM ROY, 86, visiting professor of materials, Arizona State University, died on August 26, 2010. Professor Roy was elected to NAE in 1973 "for contributions to the development of the modern science and technology of non-metallic materials."

H.E.D. SCOVIL, 86, retired director, AT&T Bell Laboratories, died on May 11, 2010. Dr. Scovil was elected to NAE in 1978 "for contributions to the conception and subsequent development for practical use of solid-state maser and magnetic bubble devices."

CLARENCE A. SYVERTSON, 84, retired director, NASA Ames Research Center, died on August 13, 2010. Mr. Syvertson was elected to NAE in 1981 "for outstanding contributions in aerospace engineering, sound guidance of research and technology programs and innovative institutional guidance."

MILTON D. VAN DYKE, 87, Professor Emeritus of Applied Mechanics, Stanford University, died on May 10, 2010. Dr. Van Dyke was elected to NAE in 1976 "for solving complex problems in aerodynamics, specifically in designs of airplanes and missiles."

Publications of Interest

The following reports have been published recently by the National Academy of Engineering or the National Research Council. Unless otherwise noted, all publications are for sale (prepaid) from the National Academies Press (NAP), 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055. For more information or to place an order, contact NAP online at *http://www.nap.edu* or by phone at (888) 624-8373. *(Note: Prices quoted are subject to change without notice. Online orders receive a 20 percent discount. Please add $4.50 for shipping and handling for the first book and $0.95 for each additional book. Add applicable sales tax or GST if you live in CA, DC, FL, MD, MO, TX, or Canada.)*

Standards for K–12 Engineering Education? The purpose of this NAE study was to assess the value and feasibility of developing and implementing content standards for engineering education at the K–12 level, similar to the content standards for the other three disciplines in STEM education—science, technology, and mathematics. To date, a small but growing number of K–12 students are being exposed to engineering-related materials, and some evidence suggests that engineering education can stimulate interest and improve learning in mathematics and science. The question is whether standards would improve the quality and lead to an increase in the teaching and learning of engineering. The study committee concludes that it would be extremely difficult to ensure the usefulness and effective implementation of content standards for several reasons: (1) we have limited experience with K–12 engineering education in U.S. elementary and secondary schools; (2) at present, there is no critical mass of teachers qualified to deliver engineering instruction; (3) evidence of the impact of standards-based educational reforms on student learning in other subjects, such as mathematics and science, is inconclusive; and (4) significant barriers would have to be overcome to introduce standards for a new content area in an already crowded curriculum.

NAE member **Robert M. White**, University Professor Emeritus, Electrical and Computer Engineering, Carnegie Mellon University, chaired the study committee. Paper, $37.75.

Technology for a Quieter America. Exposure to unwanted noise at home, at work, while traveling, and during leisure activities is a fact of life for all Americans. Very loud noise can damage hearing, and lower levels of noise can disrupt normal activities, affect sleep patterns, impede concentration at work, interfere with outdoor activities, and even cause accidents. As the world population increases and developing countries become more industrialized, problems of noise are likely to become more pervasive. Limiting exposures to noise, designing quieter buildings, products, equipment, and vehicles, and providing a regulatory environment that facilitates cost-effective, sustainable noise controls are areas that require immediate attention. This NAE report identifies the most common sources of noise and describes how they are characterized, highlights efforts to reduce noise emissions and exposures, and describes the activities of federal, state, and local agencies that regulate them for the benefit, safety, and wellness of society. The report also includes a chapter on the need for educated professionals who can deal with noise issues. This report will appeal to the engineering community; the general public; federal, state, and local governments; private industry; labor unions; and nonprofit organizations.

NAE members on the study committee were **George C. Maling Jr.** (chair), Managing Director Emeritus, Institute of Noise Control Engineering of the USA Inc., and **Richard H. Lyon**, chairman, RH Lyon Corporation. Paper, $44.00.

Investing in Dynamic Markets: Venture Capital in the Digital Age. Venture capital is a critical factor in the very existence of companies whose technical innovations created the digital revolution, and it continues to enable companies to develop new ideas into commercial products. This report goes behind the scenes of the private equity process. Drawing on his extensive personal experience, the author explains how venture capital works, why venture capitalists fund certain companies and not others, and identifies the factors that influence the success or failure of these high-risk, high-reward investments. He also

discusses the future of venture capital in a world in which the commercialization of technology requires increasingly large investments and access to global markets. Written in clear, nontechnical language, this volume features informative case studies of venture capital funding in a wide range of industries, including telecommunications, software and services, semiconductors, and the Internet.

NAE member **Henry Kressel**, managing director, Warburg Pincus LLC, co-authored the book with Thomas V. Lento. Hardcover, $37.99. Available from Cambridge University Press.

Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5. Five years ago, the National Academies report, *Rising Above the Gathering Storm*, had cautioned: "Without a renewed effort to bolster the foundations of our competitiveness, we can expect to lose our privileged position." What has changed since *Gathering Storm* was published? In the unanimous opinion of the authors of this follow-up report, the outlook has worsened. The present volume describes America at a tipping point, a point at which the federal funding necessary to address the competitiveness challenge, which will take years, if not decades, is about to be terminated. This report provides a review of what government and the private sector have accomplished in the past five years, analyzes how the original recommendations have or have not been implemented, predicts the consequences for future U.S. competitiveness, and provides a road map for moving ahead. In addition, the volume is filled with thought- and discussion-provoking factoids—many of them alarming—about the state of science and innovation in America. This urgent plea calls for sustained commitment by individual citizens and government officials at all levels. Everyone concerned about the future of innovation, competitiveness, and the standard of living in the United States will want to keep this essential tool close at hand.

NAE members on the study committee were **Norman R. Augustine** (chair), retired chairman and CEO, Lockheed Martin Corporation; **Craig R. Barrett**, retired CEO/chairman, Intel Corporation; **Charles O. Holliday Jr.**, retired chairman of the board and CEO, DuPont; **Shirley Ann Jackson**, president, Rensselaer Polytechnic Institute; **Anita K. Jones**, University Professor Emerita, School of Engineering and Applied Science, University of Virginia; **C.D. (Dan) Mote Jr.**, Glenn L. Martin Institute Professor of Engineering, A. James Clark School of Engineering, University of Maryland; **Cherry A. Murray**, dean, School of Engineering and Applied Sciences, Harvard University; **Lee R. Raymond**, former chairman and chief executive officer, Exxon Mobil Corporation; **Charles M. Vest**, president, National Academy of Engineering, and President Emeritus and professor, mechanical engineering, Massachusetts Institute of Technology; and **George M. Whitesides**, Woodford L. and Ann A. Flowers University Professor, Department of Chemistry and Chemical Biology, Harvard University. Paper, $19.95.

Advancing Aeronautical Safety: A Review of NASA's Aviation Safety-Related Research Programs. Congress asked the National Research Council to review aviation safety-related research programs of the National Aeronautics and Space Administration (NASA) and assess whether the programs have well-defined, prioritized, appropriate research objectives; whether resources have been allocated appropriately among these objectives; whether the programs are well coordinated with the safety-research programs of the Federal Aviation Administration; and whether suitable mechanisms are in place for the timely transitioning of research results into operational technologies and procedures and certification activities. Overall, the review committee concludes that NASA's aeronautics research enterprise has made, and continues to make, valuable contributions to the safety of the aviation system but has fallen short in some important respects.

NAE member **H. Norman Abramson**, retired executive vice president, Southwest Research Institute, chaired the study committee. Paper, $21.00.

Technologies and Approaches to Reducing the Fuel Consumption of Medium- and Heavy-Duty Vehicles. No fuel-consumption standards have been established for medium- and heavy-duty vehicles, such as tractor-trailers, transit buses, and trucks, which consume about 26 percent of the transportation fuel used in the United States. This report includes evaluations of technologies for improving their fuel economy and recommends approaches to developing and implementing federal regulations. The study committee recommends that regulations be based on load-specific fuel consumption, a metric that reflects the efficiency with which a vehicle moves goods or

passengers (e.g., gallons per ton-mile). The report also includes estimates of improvements that could be achieved by various technologies in the next decade for seven vehicle types.

NAE members on the study committee were **Andrew Brown Jr**. (chair), executive director and chief technologist, Innovation and Technology Office, Delphi Corporation; and **Dennis N. Assanis**, Jon R. and Beverly S. Holt Professor of Engineering, University of Michigan. Paper, $50.00.

S&T Strategies of Six Countries: Implications for the United States. Globalization has facilitated the success of formal S&T plans in many developing countries, where traditional limitations can now be overcome through the accumulation and global trade of a wide variety of goods, skills, and knowledge. As a result, centers for technological research and development (R&D) are now dispersed around the world, setting the stage for more uncertainty in the political, economic, and security arenas. This report provides analyses of the science and technology (S&T) strategies of Japan, Brazil, Russia, India, China, and Singapore, six countries that have either undergone or are undergoing remarkable growth in their S&T capabilities. The purpose of these analyses is to identify unique national features and how they have affected the global S&T environment.

NAE members on the study committee were **C.D. (Dan) Mote Jr.** (chair), Glenn L. Martin Institute Professor of Engineering, A. James Clark School of Engineering, University of Maryland; **Rakesh Agrawal**, Microsoft Technical Fellow, Microsoft Search Labs; **Robert W. Brodersen**, Emeritus Professor, Department of Electrical Engineering and Computer Science, University of California, Berkeley; **Frances S. Ligler**, U.S. Navy Senior Scientist, Center for Bio/Molecular Science & Engineering, Naval Research Laboratory; and **Fawwaz Ulaby**, Arthur Thurnau Professor of Electrical Engineering and Computer Science, University of Michigan. Paper, $35.50.

Seeing Photons: Progress and Limits of Visible and Infrared Sensor Arrays. The U.S. Department of Defense recently highlighted intelligence, surveillance, and reconnaissance (ISR) capabilities as a top priority for U.S. war fighters, and the importance of ISR assets in Iraq and Afghanistan has been widely documented in the press. The United States continues to invest in ISR capabilities, just as others continue to seek countermeasures to U.S. capabilities. The Technology Warning Division of the Defense Intelligence Agency (DIA) Defense Warning Office (DWO), which is responsible, in collaboration with other members of the intelligence community, for keeping policy makers informed about technological developments that may impact future U.S. capabilities, requested that the National Research Council (NRC) conduct a study of visible and infrared detector technologies with potential military utility that are likely to be developed in the next 10 to 15 years. This is the eighth study in a series sponsored by DWO and executed under the auspices of the NRC TIGER (Technology Insight-Gauge, Evaluate, and Review) Standing Committee.

NAE members on the study committee were **Linda P.B. Katehi**, chancellor, University of California, Davis, and **C. Kumar N. Patel**, chairman, Pranalytica Inc. Paper, $43.50.

Proceedings of a Workshop on Deterring CyberAttacks, Informing Strategies, and Developing Options for U.S. Policy. At the request of the Office of the Director of National Intelligence, the National Research Council undertook a two-phase project, a multidisciplinary investigation of strategies for (1) deterring cyberattacks on the United States and (2) the usefulness of these strategies for the U.S. government. At the end of the first phase of the study, a letter report was published that provided basic information about the nature of the problem and questions to guide research on preventing, discouraging, and inhibiting hostile activity against important U.S. information systems and networks. The second phase of the project was to select experts to address questions raised in the letter report and write commissioned papers under contract with the National Academy of Sciences. The papers were discussed at a public workshop on June 10–11, 2010, in Washington, D.C., and were revised by the authors to reflect the workshop discussions. These papers are presented as written and do not reflect consensus views of the committee. Readers should consider them points of departure for further discussion and research.

NAE member **Steven M. Bellovin**, professor, Department of Computer Science, Columbia University, was a member of the study committee. Paper, $78.50.

Nuclear Forensics: A Capability at Risk (Abbreviated Version). Nuclear forensics, the examination and evaluation of discovered or seized

nuclear materials and devices or of detonation signals or post-detonation debris from nuclear explosions or radiological dispersals, is crucial to our national security. Although the United States has developed a nuclear forensics capability that has been demonstrated in real-world incidents of interdicted materials and in exercises of post-nuclear attack scenarios, a committee of experts concludes in this new report that the U.S. Department of Homeland Security (DHS) should begin working now with cooperating agencies and national laboratories to plan and implement a sustainable, effective, appropriately funded nuclear forensics program to ensure that law enforcement and intelligence agencies can continue to use forensic evidence to prevent, mitigate, and attribute nuclear or radiological incidents. This study was requested by DHS, the National Nuclear Security Administration, and the U.S. Department of Defense.

NAE member **Milton Levenson**, consultant and retired vice president, Bechtel International, was a member of the study committee. Free PDF.

Strengthening the National Institute of Justice. Forty years ago, Congress created the National Institute of Justice (NIJ), a science agency dedicated to accumulating knowledge to support crime prevention and control by developing a wide range of techniques for dealing with offenders; identifying injustices and biases in the administration of justice; supporting basic and operational research on crime, the criminal justice system, and the involvement of the community in crime control; and awarding grants and contracts to public and private organizations and individuals. Although NIJ has gone a long way toward accomplishing its goals, its efforts have sometimes been hampered by a lack of independence, authority, and discretionary resources to carry out its mission. This assessment of the operation and quality of NIJ programs concludes that, despite its problems, NIJ is a unique organization that is vital to the nation's continuing efforts to control crime and administer justice.

NAE member **Joel S. Engel**, president, JSE Consulting, was a member of the study committee. Paper, $67.50.

Management and Effects of Coalbed Methane Produced Water in the United States. This report focuses on coal beds in the Powder River, San Juan, Raton, Piceance, and Uinta coal bed methane (CBM) basins in Montana, Wyoming, Colorado, New Mexico, and Utah. In some coal beds, natural water pressure holds methane—the main component of natural gas—fixed to coal surfaces and within the coal. When water is pumped from the coal beds, the pressure drops, and the methane is released. Water pumped from coal beds during this process—CBM-produced water—is managed through a variety of combinations of treatment, disposal, storage, or use, subject to compliance with federal and state regulations. Some states consider CBM-produced water as waste, whereas others consider it a beneficial by-product of methane production. Although current technologies can treat CBM-produced water to any desired water quality, most of it is disposed of at least cost rather than put to beneficial use because of uncertainties in the regulatory requirements. The conclusions and recommendations in this study identify gaps in data and information, potential beneficial uses of CBM-produced water and associated costs, and challenges in the existing regulatory framework.

NAE member **William L. Fisher**, professor and Leonidas T. Barrow Chair, Jackson School of Geosciences, University of Texas at Austin, was chair of the study committee. Paper, $38.25.

The
BRIDGE

(USPS 551-240)

National Academy of Engineering
2101 Constitution Avenue, N.W.
Washington, DC 20418

Periodicals
Postage
Paid

THE NATIONAL ACADEMIES™
Advisers to the Nation on Science, Engineering, and Medicine

The nation turns to the National Academies—National Academy of Sciences, National Academy of Engineering, Institute of Medicine, and National Research Council—for independent, objective advice on issues that affect people's lives worldwide.

www.national-academies.org